Equity Derivatives Markets

DISSERTATION

zur Erlangung des akademischen Grades
doctor rerum politicarum
(Doktor der Wirtschaftswissenschaft)

eingereicht an der
Wirtschaftswissenschaftlichen Fakultät
der Humboldt-Universität zu Berlin

von
Dipl.-Math. M.Sc.-Stat. Kai Detlefsen
09.01.1976, Kiel

Präsident der Humboldt-Universität zu Berlin:
Prof. Dr. Christoph Markschies

Dekan der Wirtschaftswissenschaftlichen Fakultät:
Prof. Oliver Günther, PhD

Gutachter:

1. Prof. Dr. Wolfgang Härdle
2. Prof. Dr. Rama Cont

Tag des Kolloquiums: 15. Oktober 2007

Abstract

Since the ideas of arbitrage free pricing were born, finance has changed radically - both in theory and practice. Derivatives markets have evolved and options serve nowadays as underlyings and as hedging instruments. In this thesis, we consider some markets for equity derivatives. We start by statistical analysis of the markets for European options and variance swaps because these products are important for hedging more complex claims. Then we consider different option pricing models and their calibration to observed price surfaces. Finally, we investigate the connection between option prices and the fundamental economic concept of risk aversion by the empirical pricing kernel.

Keywords:
equity derivatives, implied volatility surface, variance swap, empirical pricing kernel

Zusammenfassung

Seit der Entdeckung der arbitragefreien Bewertung hat sich das Gebiet finance grundlegend geändert - sowohl in der Theorie als auch in der Anwendung. Märkte für Derivate haben sich entwickelt und Optionen dienen heutzutage als Basis- und als Absicherungsinstrumente. In dieser Dissertation betrachten wir einige Märkte für Aktienderivate. Wir beginnen mit statistischen Analysen des Marktes für europäische Optionen und des Marktes für Varianzswaps, weil diese Produkte die hauptsächlichen Absicherungsinstrumente für komplexe Optionen sind. Dann betrachten wir verschiedene Optionspreismodelle und ihre Kalibrierung an beobachtete Preisoberflächen. Schließlich untersuchen wir die Verbindung zwischen Optionspreisen und dem grundlegenden ökonomischen Konzept der Risikoaversion anhand des empirischen Preiskernes.

Schlagwörter:
Aktienderivate, Implizierte Volatilitätsoberfläche, Varianzswap, Preiskern

Acknowledgement

Without the commitment and the energy of my advisor Prof. Dr. Wolfgang Härdle, this work would not exist in its present form. I would like to express my deep gratitude to him for constant support and numberless helpful suggestions.

Moreover, it is a great pleasure for me to thank Prof. Dr. Rama Cont for two productive and marvellous months at Ecole Polytechnique in Palaiseau where I found out about interesting new ideas in quantitative finance.

Also I am grateful for the support of Bankhaus Sal. Oppenheim. I would like to thank Dr. Peter Schwendner and the equity research group for the enjoyable internships that have inspired my research.

I gratefully acknowledge the financial support of Deutsche Forschungsgemeinschaft by Sonderforschungsbereich 649 ¨Okonomisches Risiko. Its Financial and Economic Data Center made this work possible because of the unique data and technical support.

I would also like to thank all members of the Institute for Statistics of Humboldt University Berlin for friendship and encouragement. Finally but certainly not least I would like to thank my parents, my sister and Valeria Binello for having been there whenever I needed them.

Berlin im Februar 2007 Kai Detlefsen

"Truth is much too complicated to allow anything but approximations."

John von Neumann

In memory of my mother.
In love, Kai.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation and Objectives

In the last 30 years the science of finance has changed radically undergoing an amazing growth at the same time. This change started in the 1970s when the ideas of risk-neutral valuation were published by Black and Scholes (1973) and it created huge equity derivatives markets that influence global stock markets and in this way also the world wide economy. The assets traded on derivatives markets are special contracts designed to transfer risk from parties that want to eliminate risks they face to parties that want to increase their risk exposure for adequate compensation. Hence, derivatives markets help market participants to create the risk-return profiles they want to have. This possibility explains also why derivative markets exist in addition to the markets for the underlying instruments. Derivatives are used by different groups, including corporations, hedge funds and financial institutions, for different reasons. Hedgers want to insure their portfolios against unfavourable movements of assets by buying special derivatives on these assets. But there are also speculators who take directional views on the market and try to benefit from the leverage effect of derivatives. In the last decade, the exchange-traded derivative markets have grown rapidly. In this segment contracts on indices have the biggest market share. Because of the simple structure and high liquidity these standard products are often used for hedging more complex derivatives that are sold over the counter. These so-called exotic options are developed to meet particular needs of clients. Because of their special structure these products are difficult to hedge by standard options.

The pricing and hedging of complex options are carried out in stochastic models using the idea of risk-neutral evaluation. This concept was introduced

by Black and Scholes (1973) who won the Nobel Prize for economics in 1997 for their discovery. Their seminal principle is a cornerstone in the theory of modern finance and at the same time it is applied by traders every day in industry. The idea determines the price of an option by the cost of setting up a perfect hedging strategy. In the Black-Scholes model all options can be replicated perfectly by continuously rebalancing a self-financing portfolio of stocks and bonds. This pricing by replication is equivalent to computing an expected value under a so-called risk-neutral measure. This approach to compute accurate option prices made finance a hard science like the natural science and spurred the development of the derivatives markets.

The Black-Scholes model is used nowadays rarely for option pricing because empirical analyses show significant differences between model and market prices. But nonetheless the Black-Scholes option pricing formula is applied in industry for quoting the prices of European options. These prices are quoted in implied volatility that is defined as the number that makes the Black-Scholes price and the market price equal if the number is used as volatility. In the Black-Scholes model, the volatility is assumed to be constant but in reality the implied volatility changes with strike level and time to maturity. The surface that describes the implied volatility as a function of strike and time to maturity is commonly referred to as the implied volatility surface. It has been the object of investigation in numerous studies because it reflects basically the prices of the liquid market of European puts and calls.

Because of these shortcomings of the Black-Scholes model other modelling approaches have been analyzed. Many of these models propose a more sophisticated stochastic model for the evolution of the price process of the underlying asset. But the risk neutral valuation technique that Black and Scholes motivated by replication is still applied in almost every model so that it can really be regarded as a cornerstone of finance. One strand of literature keeps the Black-Scholes model and tries to explain the implied volatility surfaces by market frictions like transaction costs, illiquidity or other trading restrictions. Figlewski (1989) analyzed the effects of transaction costs and concluded that they can explain to a great extent the form of the implied volatility surface. In equity markets transaction costs are not negligible and have to be taken into account for hedging and hence also for pricing. But the explanation of the implied volatility surface by trading restrictions is difficult because many effects interact so that the overall effect is hard to quantify. In general, these practical issues are regarded only as a partial explanation.

Academia and industry focus instead on other modeling approach that allow a quantification. As the implied volatility surface shows that the volatility depends on strike and time to maturity one strand of literature relaxes the Black-Scholes assumption of a constant volatility: Derman and Kani

(1994a), Derman and Kani (1994b), Dupire (1994) and Rubinstein (1994) showed that there exists a unique deterministic function of time and stock price such that the Black-Scholes model with this local volatility function exactly replicates observed implied volatility surfaces. As the local volatility function is nonparametric these models give a perfect in-sample fit but they often lead to unrealistic dynamics of implied volatility surfaces. A mathematically more involved approach are the stochastic volatility models where the instantaneous volatility is modeled by another stochastic process that may be correlated with the stock price process. Such models have been proposed by Hull and White (1987), Scott (1987) and Heston (1993). A randomly changing volatility is economically appealing because volatility clustering is a stylized fact of equity returns. Moreover, the out of sample performance of these models is more realistic than in the local volatility models. But a stable estimation of the model parameters from implied volatility surfaces turns out to be difficult as it often leads to regularized nonlinear constraint optimization problems. Besides these approaches that change the volatility, there are also attempts to use fat-tailed distributions for the innovations of the returns instead of the normal innovations. These models describe the stock price process by stochastic differential equations that are not driven anymore only by Brownian motions. An example of this model class gave Merton (1976) by adding jumps to the returns in the Black-Scholes model. Jumps reflect stock market crashes and hence have attracted considerable attention recently. These approaches can also be combined in a stochastic volatility model with jumps in the stock returns as Bates (1996) showed. The deterministic and stochastic volatility models are all calibrated to the implied volatility surface and try to replicate these option prices by the stock price process. Another type of models takes the liquid market of European options as given and models these prices as implied volatilities by stochastic differential equations. This method is justified economically because the market's high liquidity ensures fair prices so that these options do not have to be priced theoretically any more. The difficulty of this approach lies in guaranteeing that the option price processes described in implied volatility are martingales. Schönbucher (1999) considered this problem for the term structure of volatility and Ledoit et al. (2002) for the whole implied volatility surface. Empirical aspects of this approach have been analyzed by Cont and da Fonseca (2002) who identify factors driving the surfaces. As this approach is quite complex for implied volatility surfaces Bühler (2006) considered a similar method by taking variance swap curves in addition to the stock price as primitive assets.

Besides these microeconomic models asset pricing is also considered in macroeconomics by equilibrium models. This approach requires the spec-

ification and estimation of many economic concepts like utility functions. Hence it is not appropriate for the quantitative problems the industry faces. But it gives rich qualitative answers to fundamental economic questions. The combination of these two asset pricing approaches allows to derive interesting conclusions by the pricing kernel. Jackwerth and Rubinstein (1996) analyzed in this way the market utility function that is given by the integral of the empirical pricing kernel. He identified in the empirical utility function a region where the representative investor behaves risk seeking. This leads to questions of behaviourial finance that Kahneman and Tversky (1979) analyzed. They received the Nobel Prize in economics for their work that developes a new theoretical basis for modelling decisions of investors and hence lies at the heart of finance. These two Nobel Prizes have motivated this work to analyze option markets in the risk neutral world and to derive the implications about the behaviour of investors.

The overall aim of this work is to analyze option markets empirically and to investigate models using real data. Moreover, we consider as application the fundamental economic question what behaviour of investors is implied by equity derivatives prices. This question can be answered by comparing the risk neutral and the real world in a stochastic sense.

1.2 Structure

This thesis is organized in the following way: In chapter 2, we take an overview on derivatives markets and see how they evolved and changed in time. Then we analyze stylized statistical facts about the returns of the DAX because we focus in the following mainly on options with the DAX as underlying asset. After the analysis of the underlying we consider the price dynamics of European options on the DAX with long times to maturity. Using a principal components analysis we identify typical changes of the implied volatility term structure – in contrast to earlier works that found the strike dimension more important for the modelling. Besides European options, variance swaps have become quite popular and their markets are also very liquid having sometimes half the volume of the corresponding market of European options. These swaps are mainly used for speculating on realized volatility and for hedging complex options on realized volatility. Because of this importance we analyze the market of variance swap for the SPX using an option pricing model, a parametric descriptive approach and a semiparametric model. In an out-of-sample analysis we conclude that variance swap curves can be forecasted well for long times to maturity.

In chapter 3, we give an overview on current modelling approaches for

equity derivatives. In particular, we present a semiparametric stochastic volatility model that perfectly replicates variance swap curves. This model can be regarded as a combination of the nonparametric local volatility model of Dupire (1994) and the parametric stochastic volatility model of Heston (1993). Moreover, we consider different valuation techniques for derivatives and discuss some implementation details.

In chapter 4, we discuss the calibration of option pricing models to observed price surfaces. A stable solution of this problem is essential for applications like pricing or hedging of exotic options. We consider the calibration of the objective measure from a time series of the underlying and the estimation of the risk neutral measure from option prices. In particular, we analyze different specifications for measuring the error between the market and the model prices. Finally, we examine different ways to regularize the problem in order to stabilize the estimation results.

In chapter 5, we consider an applications of option pricing models by analyzing a question that led to the Nobel prize for Kahneman and Tversky in 2003: We estimate empirical pricing kernels from DAX data and option prices. The resulting utility functions show a region near the initial wealth where the market investor behaves risk seeking. This untypical form of the utility function is explained by considering individual investors. The distribution of individual investors who are assumed to have a generalized Kahneman utility function is given by an inverse problem. Considering three different market regimes we analyze how the individuals' behaviour changes over time and markets.

Chapter 2

Equity Derivatives Markets

In this chapter, we introduce equity derivatives markets, consider statistical properties of typical underlying instruments and analyze the fundamental markets of European options and variance swaps. Section 2.1 describes derivatives markets in general, their evolution over time and then focuses on equity derivatives markets. In section 2.2, we analyze empirical properties of returns of a stock index because such indices are the underlying asset for many derivatives. The market of European call and put options is modelled in section 2.3 by non-parametric factors. In section 2.4, we study different modelling approaches for the market of variance swaps that also serve as hedging instruments nowadays.

2.1 Derivatives Markets

Assets such as stocks and bonds have been traded for a long time on exchanges. These markets are of paramount importance for the economy because they bring together investors looking for income and corporations in need of capital. In addition to these primary assets, exchanges trade also products whose values depend on other underlying instruments. In this section, we describe these derivative markets.

Derivatives are not only traded on exchanges, also financial institutions buy and sell directly contingent claims. This market that is formed by banks and their clients is called the over-the-counter market. Its advantage is the freedom to trade any contract and not only those specified by the exchanges. A price to pay for this freedom is the risk that a party in an over-the-counter trade cannot fulfill its obligations. Exchanges on the other hand have almost eliminated such credit risk.

The over-the-counter and the exchange-traded market for derivatives form

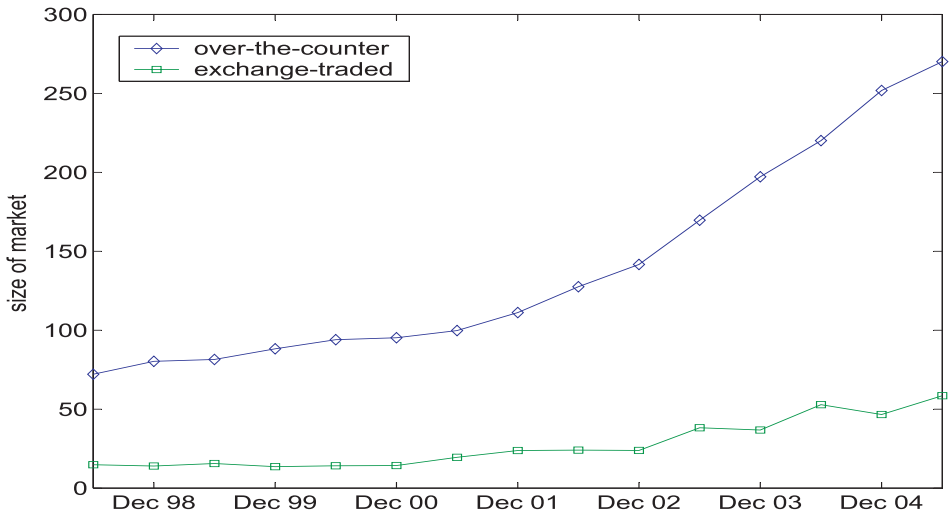

Figure 2.1: Size of over-the-counter and exchange-traded derivatives markets. (in trillions of US dollar)

together the derivatives market. Its products do not only serve for investing or speculating but they can be also used for reducing and controlling the risk of portfolios. The importance of this way to control risk is reflected in the volume and continuous growth of the derivatives markets. Figure 2.1 shows the estimated total amounts underlying transactions that were outstanding in both markets between 1998 and 2005. The data have been collected by the Bank for International Settlement. Both markets show a clear and stable trend upward. Moreover, the over-the-counter market has always had a higher volume than the exchange-traded market and this difference has increased in absolute and relative terms over the years. In June 2005, the over-the-counter market grew to $ 270 trillion while the exchange-traded market had a volume of $ 52 trillion. These huge volumes are the principal amounts that were underlying the outstanding transactions.

The values of the outstanding contracts were much lower. These gross market values are presented in figure 2.2. While the size of the over-the-counter market was $ 270 trillion in June 2005 the corresponding gross size was less than $ 11 trillion, hence the gross size of the market is estimated as 4 % of the amounts underlying the transactions. Moreover, figure 2.2 shows that the gross size also grew over the years in the mean but there were years when it decreased compared to the preceding year. In addition, this figure presents the evolution of the gross size of the individual over-the-counter markets, i.e. foreign exchange, interest rate, equity and commodity

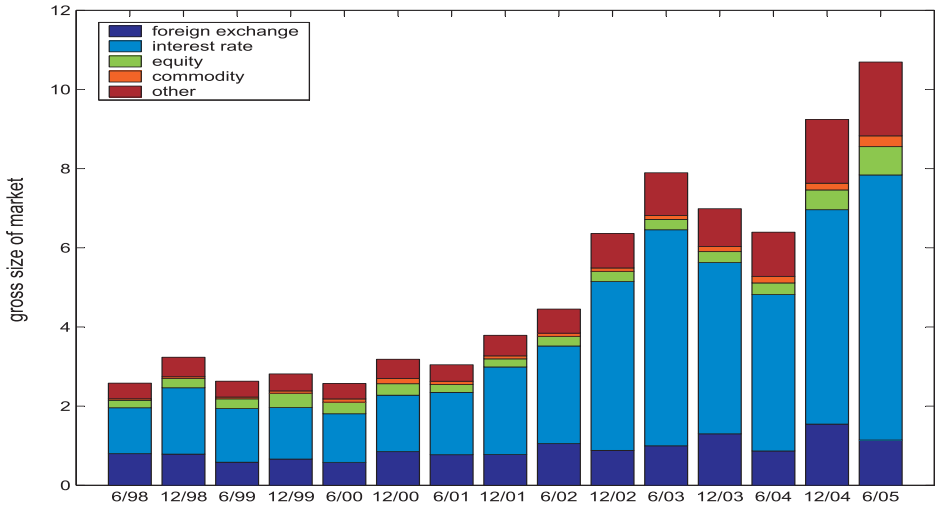

Figure 2.2: Gross size of over-the-counter derivatives markets. (in trillions of US dollar)

markets. While the gross size of the foreign exchange market was more or less constant in absolute terms, all other markets have grown: the gross size of the interest rate market increased between 1998 and 2005 by the factor 5.8 and the equity market grew over this time period by the factor 3.8. The interest rate market clearly has become the biggest over-the-counter market and the equity market has almost attained the size of the foreign exchange market. We concentrate in this work on the equity derivates market that grew in the last years the most in relative terms and will probably become the second biggest over-the-counter market in the next years.

The equity market of over-the-counter derivatives can be divided into options and forward contracts. The growth of these two market segments is presented in table 2.1. The option segment always had a much bigger gross size but both segments grew over the years in a stable way. The option and forward markets on European and US equities constitute almost the whole market and the products on European equities had in many years the highest gross size.

Although the exchange-traded equity market is much smaller than the over-the-counter market it is fundamental for the whole derivatives business because over-the-counter products are often hedged by options of the exchange-traded market. Hence, we analyze the liquid market of European options in section 2.3 after a statistical analysis of a typical underlying. In section 2.4 we consider the variance swap market that is also important for

Market	6/98	6/99	6/00	6/01	6/02	6/03	6/04	6/05
Forwards and Swaps								
US equities	8	35	21	18	22	22	23	26
European equities	9	12	29	23	35	39	32	50
Japanese equities	2	2	6	3	1	1	2	2
Other equities	2	3	6	4	4	3	7	11
Options								
US equities	42	48	50	38	42	54	71	74
European equities	107	129	161	100	123	123	132	180
Japanese equities	4	8	10	5	7	6	13	20
Other equities	17	8	10	8	9	10	15	18

Table 2.1: Gross market values of OTC equity derivatives by market and instrument. (in billions of US dollar)

hedging purposes.

2.2 Underlying assets

Most equity derivatives pricing approaches are based on models of the price process of the underlying asset. These models are often built in order to replicate special properties of the derivatives markets or in order to allow simple numerical pricing. But at the same time, they should reflect stylized facts of the underlying price process. In this section, we discuss some typical features of equity price processes and see what they imply for the modelling.

In finance, we often do not consider directly the price process (S_t) but focus instead on the returns $r_t := (S_{t+\Delta} - S_t)/S_t$ over periods of length Δ. Alternatively, log returns $R_t := \log S_{t+\Delta} - \log S_t$ can be analyzed which are good approximations to returns r. Actually, the returns are the first order expansions of the log returns.

An important issue for the modelling of returns and hence of asset prices is the degree and type of dependence between the returns. The past behaviour of returns does not necessarily reflect their future performance. This economic statement can be interpreted in a statistical way that the autocorrelations of the returns are not significantly different from 0. In figure 2.3, we present the autocorrelation function for DAX returns with approximate 95% confidence bands. This figure shows that there is no linear relationship between the returns. But if the time scale is changed then significant auto-

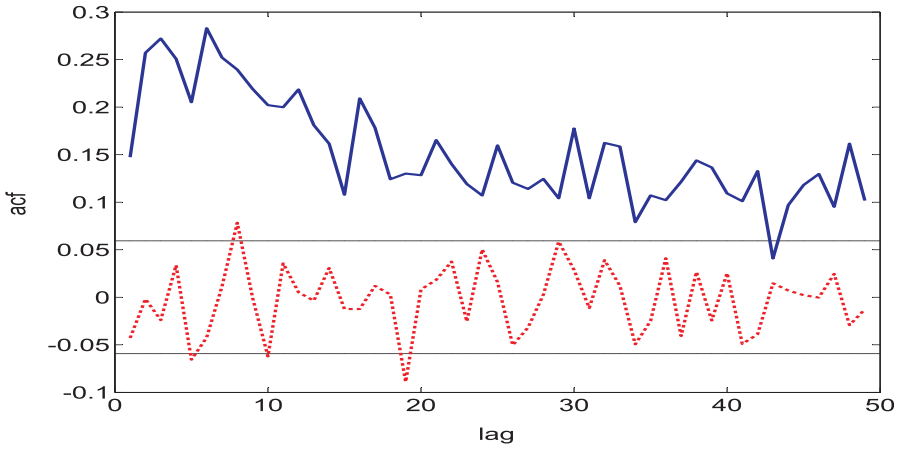

Figure 2.3: Autocorrelation of daily DAX returns (red, dotted) and squared returns (blue, solid), 01/2000 - 06/2004.

correlations can exist. For very short time scales they can be attributed to the market micro structure and for long time scales we have less data so that the evidence can be seen as less conclusive.

But returns are not independent because nonlinear transformations of returns often manifest positive autocorrelation. This relationship is also illustrated in figure 2.3 where we consider the squared returns. The positive correlation implies that a high squared return is often followed by a high squared return. Hence, prices tend to build regimes of high variation and of low variation. This phenomenon is called volatility clustering. It is also illustrated in figure 2.4 that shows daily DAX returns. The changing volatility is visible, e.g. at the end of 2002 and at the beginning of 2003 the volatility is high while the year 2000 is a low volatility regime. This shows that returns should not be modelled as independent random variables and volatility may not be constant.

Another stylized fact of returns is the leverage effect: Negative returns are often followed or accompanied by a higher volatility. This phenomenon can be quantified by the correlation between returns and subsequent squared returns. These correlations are shown in figure 2.5 as a function of the time between returns and squared returns. The negative correlation confirms the adverse movements of prices and volatility. This dependence vanishes if the time between the returns and squared returns increases. Moreover, we present in figure 2.6 how the 1 day leverage effect changes in time estimated from the last 60 days. Despite the high variation we see that the correlation is negative in general.

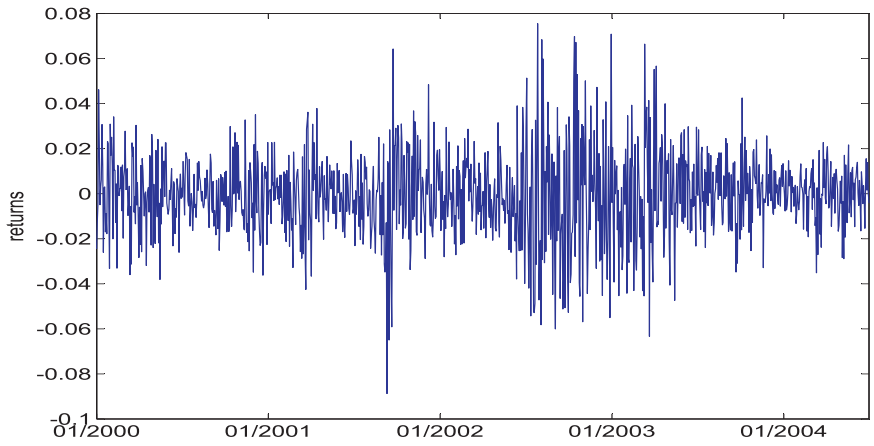

Figure 2.4: Daily DAX returns, 01/2000 - 06/2004.

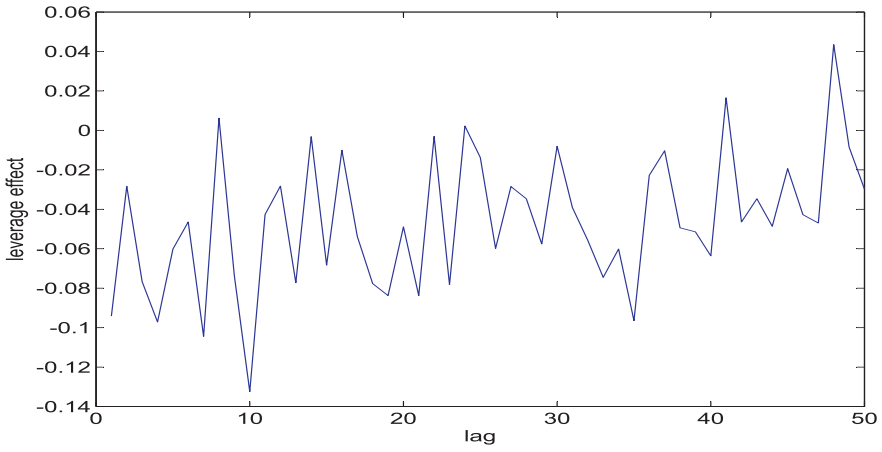

Figure 2.5: Leverage effect in DAX returns (01/2000 - 06/2004) measured by correlation between lagged returns and squared returns.

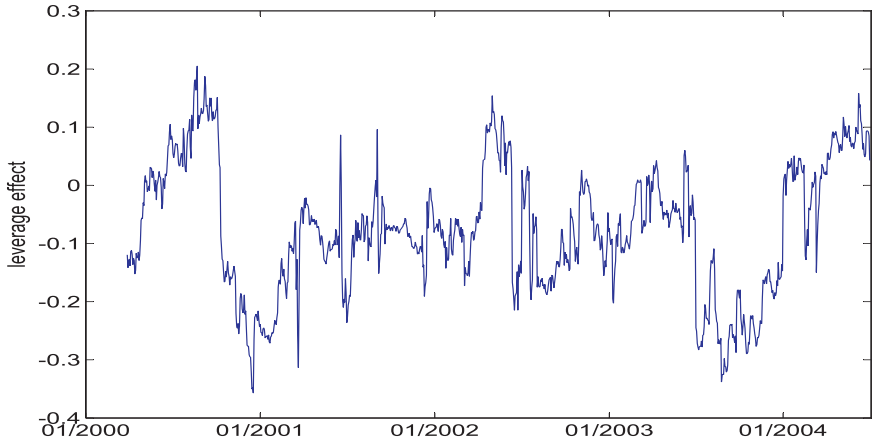

Figure 2.6: 1 day leverage effect in DAX returns, 01/2000 - 06/2004.

In equity derivatives pricing, the drift of the underlying price process is not important because in a risk-free world the drift is fixed, i.e. depends on the interest rate. Hence, we do not discuss questions of stationarity or existence of local trends in the prices. But we consider the dispersion of the returns which is closely connected to volatility - a central concept for option pricing. To this end, we present in figure 2.7 a kernel density estimate of the DAX returns. (We use a normal kernel and choose the bandwidth by Silverman's rule of thumb.) As the Black-Scholes model uses normally distributed returns we show in this figure also a kernel density estimate of a normal random variable with the same mean and variance. Even without confidence bands we see clearly that the returns do not follow a normal distribution. Besides the higher peak, the returns also have fatter tails as is shown in figure 2.8. A number of distributions have been proposed for modelling these features. Although none of these distributions has prevailed it seems clear that returns are not normally distributed.

Besides the distributional modelling of the price process, the properties of the paths are also important. In this context, the conclusions depend a lot on the time scale at which the price process is observed. At a micro level, the path are of course discontinuous. The longer the periods between the observations are the less apparent become the jumps in the price process. Considering a daily time series of DAX prices it is not clear if there exist jumps in the process. In order to answer it we analyze the returns. While a jump can be defined by an economically motivated threshold level we apply a statistic test for outlier detection. We use Grubb's test which is based on a normality assumption that can be accepted as approximation for the consid-

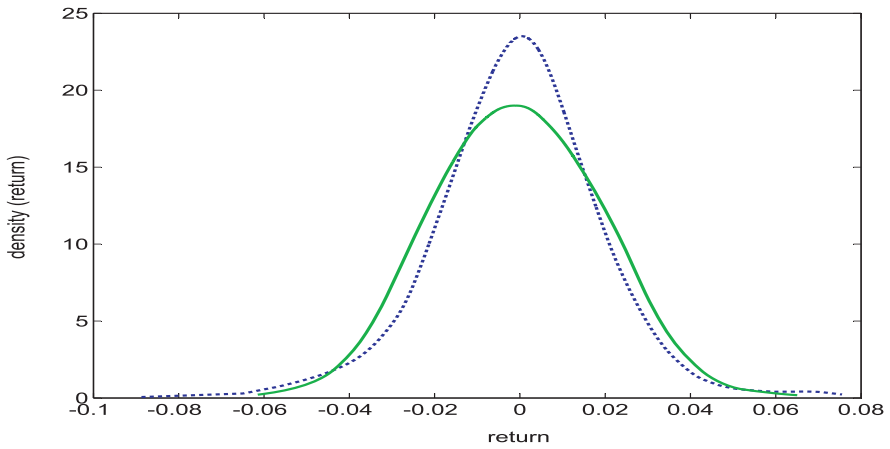

Figure 2.7: Kernel densities of observed returns (blue, dotted) and samples of a normal distribution with the same mean and variance (green, solid).

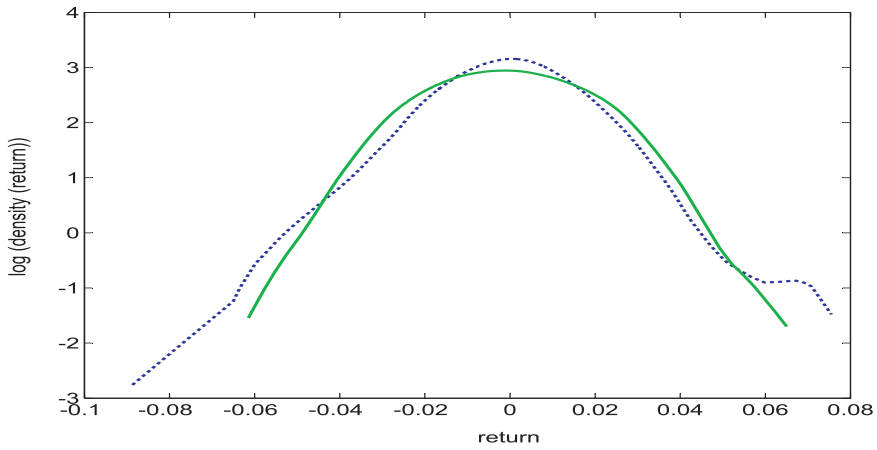

Figure 2.8: Logarithm of kernel densities of observed returns (blue, dotted) and samples of a normal distribution with the same mean and variance (green, solid).

ered data. The one-sided test statistic $(\bar{r} - r_{min})/s_r$ has under assumption that r_{min} is no outlier a distribution that can be characterized in terms of the t-distribution. Applying also the analog test for the maximum we identify two outliers: A negative return of 8.8% on 9/11/2001 and a positive return of 7.6% on 7/26/2002. These unusual returns can be attributed to political events. This discussion about jumps in asset prices can be summarized in such a way that the modelling of the prices can but need not incorporate jump components.

In addition to stochastic jumps, the price process also jumps because of dividend payments. It may jump when the dividend is announced and will fall by the dividend amount when the asset goes ex-dividend. On the other hand, derivatives prices do not change on the ex-dividend date because the payment is known. But these prices can jump when an unexpected dividend is announced. Because of market frictions, these effects differ sometimes in reality from theoretical considerations. E.g. taxation in different countries can have different impacts on a dividend of the EuroStoxx50. In order to model the uncertainties of dividends an additional stochastic factor can be incorporated into the model. As we consider mainly the DAX that is a performance index we do not consider or model dividends explicitly.

2.3 Market for European Options

The markets of European index options are of paramount importance for option pricing because models are calibrated to the prices of these options that serve as hedging instruments for more complex products. Hence, we analyze statistically the movements of the price surfaces of DAX options focusing on long times to maturity that are traded since a short time ago. Via a principal components analysis we identify the typical movements in the term structure and not in the smile.

As the components have only limited explanatory power we determine classes of typical movements by a cluster analysis and show how much variation is explained by the clusters and principal components in each cluster. As the overall variation can be well explained by two components per cluster we propose a model for the dynamics of the surfaces by modelling the switching between the clusters by a Markov chain. This Markov chain is estimated and the estimation of the time series of factor loadings is also discussed. The resulting model generalizes the direct principal components approach.

2.3.1 Introduction

Implied volatility surfaces represent the prices of European call and put options. By inversion of the Black-Scholes formula every price of such an option with maturity T and strike K implies a volatility $\sigma(T, K)$ that replicates the price under the assumptions of the Black-Scholes model. Option pricing models are calibrated to these surfaces σ and then exotic options are priced with the estimated models. As exotic options are often hedged with European options it is essential for a correct pricing of exotic options that the model replicates the implied volatility surfaces in order to assess the hedging costs correctly. Moreover, some modern derivatives require a significant rebalancing of the hedge with European options over time. For such options it is in addition important that the model makes reasonable forecasts of the implied volatility surfaces. Thus, understanding these surfaces and their dynamics is fundamental for pricing and hedging exotic options.

Implied volatility surfaces show some characteristic patterns over time like the "smile" in the strike dimension and its flattening in the maturity dimension. But the surfaces change over time and we analyze these movements of the surfaces by a principal components approach. Recently, exchanges started to trade European index options expiring in several years. As these derivatives are used for hedging exotic options with corresponding times to maturity our analysis of implied volatility surfaces comprises these new and important price information of options expiring in several years. The analysis of these long surfaces identifies – in contrast to earlier studies – time to maturity as the important dimension.

By a principal components analysis we identify factors that drive the evolution of the DAX surfaces. In contrast to the analysis of Cont and da Fonseca (2002) that consider only implied volatility surfaces with short times to maturity we identify the major movements in the term structure. Moreover, we conclude that a high explained variance requires more factors than in Cont's analysis. As only the first factors can be interpreted economically the resulting factor model either explains only a part of the variation or contains also purely statistical factors. Hence, we see difficulties in this approach for long implied volatility surfaces. Because of this, we consider the movements of the surfaces in a cluster analysis. The resulting groups identify typical movements. Performing principal components analysis in each cluster we see how much variation is explained by the clustering and by the components. As already a few components give a good explanation we propose a model that switches between the clusters by a Markov chain. We estimate the transition matrix and explain how to estimate the processes for the factor loadings that can be observed only when the Markov chain is in the cluster. The proposed

model generalizes the approach of Cont and da Fonseca (2002) who model all movements by the same components.

Recently, Fengler (2005) analyzed DAX implied volatility surfaces with short times to maturity in a dynamic semiparametric model. In this model, the factors are estimated non-parametrically and combined in a linear way. Fengler identified three factors that explain almost all of the variation where the first factor accounts already for 95% of the variation. The factors can be interpreted as level, slope and curvature in the strike dimension. This modelling approach does not work satisfactorily for DAX surfaces with long times to maturity because there are in general not enough observations with long times to maturity for a stable calibration of the model. Cont and da Fonseca (2002) also identified three factors that move SPX and FTSE surfaces with short and medium maturities. In their functional principal components analysis, the resulting factors have the same interpretation as level, slope and curvature in the strike dimension.

This work is organized as follows: In Section 2.3.2 we introduce implied volatility surfaces, describe the DAX data and their smoothing. In Section 2.3.3 we identify and interpret the estimated factors after a description of the principal components method. In Section 2.3.4 we estimate and discuss typical regimes by a cluster analysis. Then we propose for the movements of the implied volatility surfaces a model based on a Markov chain that switches between these clusters. In Section 2.3.5 we summarize the results of this empirical analysis of DAX implied volatility surfaces with long times to maturity.

2.3.2 Implied volatilities

A European call on an asset S_t with strike K and maturity T offers the option holder the right to buy the asset at time T for the strike price K. Hence, the payoff of this option is $(S_T - K)^+$. In the Black-Scholes model this option can be replicated perfectly and its price $C_t(K,T)$ at time t is given by the Black-Scholes formula

$$C_t(K,T) = S_t\Phi(d_+) - Ke^{-r(T-t)}\Phi(d_-)$$

where $d_\pm = (-\log m + (T-t)(r \pm 0.5\sigma^2))/(\sigma\sqrt{T-t})$. Here, $m = K/S_t$ denotes moneyness and Φ is the cumulative distribution function of the normal distribution.

Given a market price $C_t^M(K,T)$ of a call we can find a volatility $\sigma_t(K,T)$ such that the market and the model price coincide if this volatility is used in the Black-Scholes pricing formula: $C_t(K,T) = C_t^M(K,T)$. Existence and

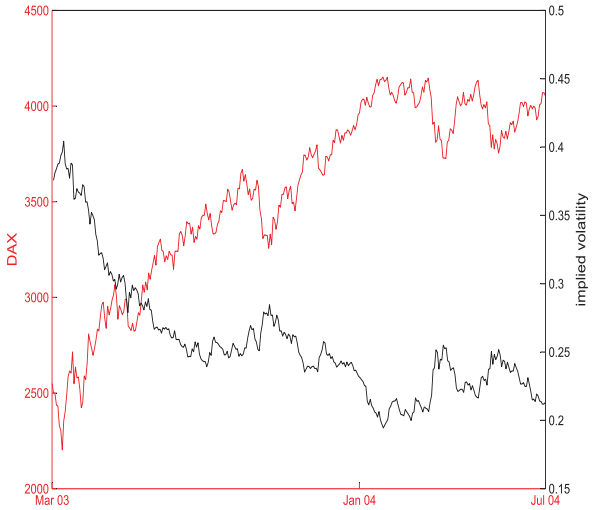

Figure 2.9: DAX and implied volatility at the money with 1 year to maturity, March 2003 - June 2004.

unique of this volatility follow from the monotonicity of the Black-Scholes price as a function of volatility.

Hence, we can compute at time t for every pair of strike K and maturity T a volatility implied by the corresponding market price. In this way, we observe every day t an implied volatility surface $(K,T) \mapsto \sigma_t(K,T)$. In order to make these surfaces comparable on different days one often uses moneyness $m = K/S_t$ instead of strike and time to maturity $\tau = T - t$ instead of maturity. If we denote implied volatility surfaces in these coordinates by I_t then we have $I_t(m,\tau) = \sigma_t(mS_t, t+\tau)$.

We analyze closing prices of European call and put options on the DAX, the German stock index. The prices are represented each day by an implied volatility surface. Before March 2003 there were only index options with short and medium times to maturity but since March 2003 European calls and puts with maturities up to 5 years are traded. Hence, we analyze the time series of DAX implied volatility surfaces from March 2003 to June 2004. A general description of the market in this period is provided by figure 2.9 that shows the evolution of the DAX and the implied volatility at the money for 1 year to maturity.

Every day t we observe the implied volatilities $I_t(m_i,\tau_i)$ of the traded options $(i = 1,\ldots,n)$ with moneyness m_i and time to maturity τ_i. The number of options n ranges from 170 to 323 and is in the mean 261. The

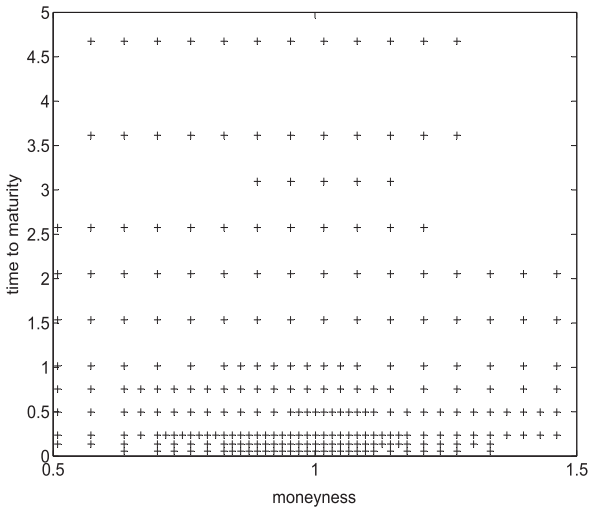

Figure 2.10: Moneyness/maturity points of option prices on 01 June 2003 (only for moneyness between 0.5 and 1.5).

distribution of moneyness and time to maturity is presented in figure 2.10 on a typical day. For each time to maturity the moneyness grid is regular with more observations for short maturities than for long maturities. Moreover, we see that there are more options with short times to maturity. This grid changes every day because the time to maturity of all options decreases and not all options are traded every day.

As the moneyness range of observed prices with long times to maturity is limited we analyze in the following only options with moneyness between 0.75 and 1.25. In the maturity dimension we consider options with maturity up to 4 years. For the principal components and cluster analysis we smooth the surfaces on a fixed grid. In the moneyness direction, we use an equally spaced grid: $m = 0.75, 0.8, \ldots, 1.25$. In the time to maturity dimension, we use a grid of 4 months: $\tau = 0.33, 0.67, 1, 1.33, 1.67, 2, 2.33, 2.67, 3, 3.33, 3.67, 4$.

The smoothing of implied volatility surfaces is done in a nonparametric way. We do not consider no-arbitrage constraints as there are different ways of smoothing under these difficult constraints. We determine the surfaces on the fixed grid by the Nadaraya-Watson estimator:

$$I_t(m,\tau) = \frac{\sum_{i=1}^n \mathcal{K}_{h_1,h_2}(m-m_i,\tau-\tau_i) I_t(m_i,\tau_i)}{\sum_{i=1}^n \mathcal{K}_{h_1,h_2}(m-m_i,\tau-\tau_i)}$$

where $\mathcal{K}_{h_1,h_2}(m,\tau) = \exp(-m^2/(2h_1))\exp(-\tau^2/(2h_2))/(2\pi)$ is the product

of two Gaussian kernels.

The choice of the bandwidths h_1 and h_2 is essential in smoothing because it represents a trade-off between oversmoothing and overfitting: The higher the bandwidths the smoother is the resulting surface and the bigger is the bias. On the other hand, small bandwidths lead to a high variation in the smoothed surface with a small bias. These parameters can be determined globally by cross validation, see e.g. Härdle et al. (2004). Another method are adaptive bandwidth estimators, see e.g. Gasser et al. (1991). Instead of an automatic bandwidth selection we choose local bandwidths that are motivated by the structure of the observations, see figure 2.10. In the moneyness direction the grid has a regular design where the distance between the observations is smaller for smaller times to maturity. Thus, we use a corresponding bandwidth h_1 proportional to this distance. In the maturity direction the grid changes every day, hence we estimate the density of the times to maturity by a local constant regression and choose the corresponding bandwidth h_2 proportional to this density.

Figure 2.11 shows the average implied volatility surface of the time series of smoothed surfaces. This surface decreases in moneyness direction for all times to maturity. Hence, we observe clearly the skew of the implied volatility surfaces. In the time to maturity dimension the implied volatilities increase for big moneyness and decrease for small moneyness. At the money the average surface shows a relatively constant term structure. Moreover, we present in figure 2.11 the standard deviation of the implied volatility surfaces. This plot shows that the surfaces fluctuate more for short times to maturity.

2.3.3 Principal components

In this section, we review an approach of Cont and da Fonseca (2002) that describes the movements of implied volatility surfaces by factors. First, we describe the underlying statistical method, principal components analysis.

We observe on each day t a surface that is represented by a matrix of implied volatilities $(I_t(m,\tau))$, $m = 0.75,\ldots,1.25$, $\tau = 0.33,\ldots,4.0$. We interpret this matrix as a vector $X = (X_1,\ldots,X_p)$, $p = 132$. In principal components analysis one looks for a weighted average of this vector

$$\delta^\top X = \sum_{j=1}^{p} \delta_j X_j$$

such that $\sum \delta_j^2 = 1$. The weighting vector δ is chosen such that the variance of the projection $\delta^\top X$ is maximized

$$\max_{\{\delta:\|\delta\|=1\}} \mathrm{Var}(\delta^\top X).$$

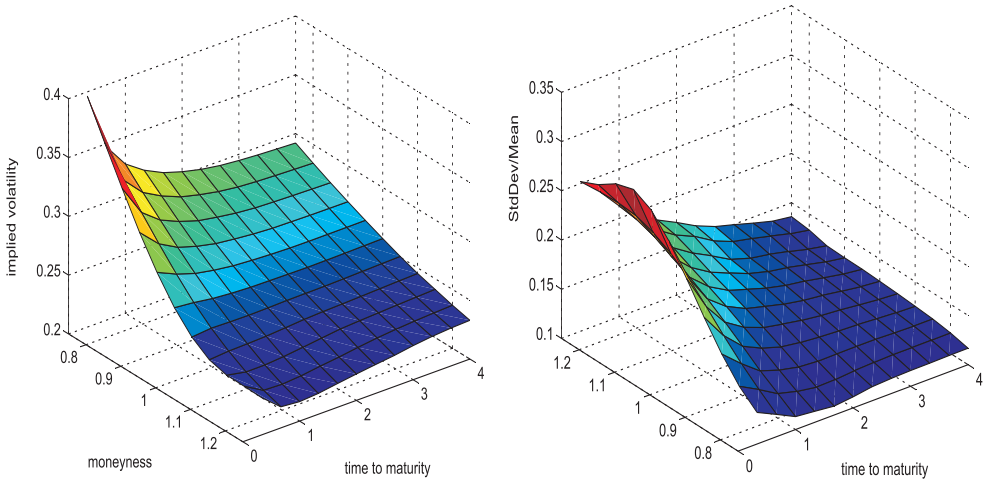

Figure 2.11: Average implied volatility surface (left) and daily standard deviation/mean of implied volatility surfaces, Mar 2003 - Jun 2004.

The quadratic form is maximized by the eigenvector γ corresponding to the largest eigenvalue λ of the covariance matrix $\mathrm{Var}(X)$. This eigenvector is called the first principal component. Proceeding in this way we see that the principal components are given by $\Gamma^\top X$ if $\mathrm{Var}(X) = \Gamma\Lambda\Gamma^\top$. When applied to data corresponding estimators have to be used e.g. for the covariance matrix.

As principal components analysis is sensitive to scale changes it should only be applied directly to data of the same scale. In our problem of implied volatility surfaces the variables all have the same scale so that we can apply the technique directly. How well the first q principal components describe the data can be measured by the proportion of explained variation of total variation. This is given by $\sum_{i=1}^{q} \lambda_i / \sum_{i=1}^{p} \lambda_i$ where λ_i is the eigenvalue corresponding to the ith eigenvector. For more details on this technique, we refer to Härdle and Simar (2003). Functional principal components analysis, the continuous analog of the described discrete method, leads to similar results because implied volatility surfaces and their movements are rather smooth.

We do not determine the principal components of the surfaces directly but consider instead the difference of the log transformed data

$$\Delta X_t(m,\tau) = \log I_t(m,\tau) - \log I_{t-1}(m,\tau).$$

Hence, we consider relative movements of the surfaces. This transformation that was also used by Cont and da Fonseca (2002) ensures positivity of

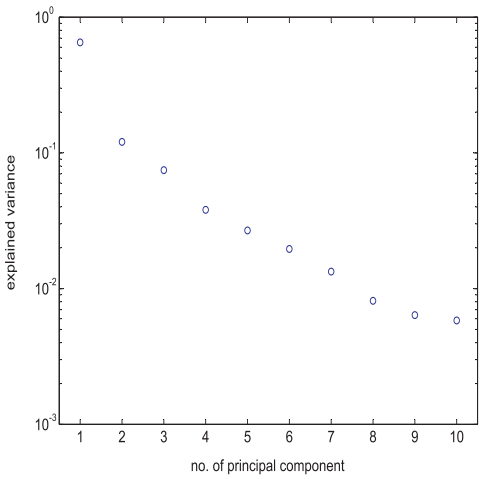

Figure 2.12: Relative proportion of variance explained by principal components.

principal component	1	2	3	4	5	6
cum. expl. variance	65.2	77.3	84.5	88.5	91.2	93.2

Table 2.2: Cumulative explained variance.

the surfaces and makes the analyzed surfaces more stationary. The original surfaces can be recovered by

$$I_t(m,\tau) = I_0(m,\tau)\exp\left\{\sum_{i=1}^{t}\Delta X_t(m,\tau)\right\}. \tag{2.1}$$

We apply a principal components analysis to the surfaces ΔX_t. The results of the principal components analysis are given by the factors, their loadings and the variance the factors explain. In figure 2.12, we present how much of the total variation is explained by the factors. The first factor accounts for 65% of the variation, the next for 12%. Each of the first five factors accounts for more than 2%, the factors after the 7th all explain less than 1%. In table 2.2, the cumulative explained variance is shown. The first three factors explain 84% and the first five 91%. Hence, the explained variance is smaller than in Cont and da Fonseca (2002) who analyzed shorter surfaces. The first six principal components are shown in figure 2.13. The first factor represents the level because the whole surface shifts up if the factor has a positive weight. Moreover, it changes the slope of the term structure because the factor is bigger for shorter times to maturity. The second factor

represents only the slope of the term structure because it is positive for short and negative for long times to maturity. The factor three to six are harder to interpret but represent in general the term structure and not the smile. This stands in contrast to the meaning of the factors in Cont and da Fonseca (2002).

A model for implied volatilities can be derived from equation (2.1) by modelling the time series of factor loadings. These series could be modelled e.g. by univariate AR(1) processes as in Cont and da Fonseca (2002). This modelling of all implied volatility surface movements by a principal components analysis leads to good results for small surfaces. But for long surfaces the principal components have short explanatory power and the form of the factors becomes hard to interpret.

2.3.4 Cluster analysis

The direct principal components analysis leads to a satisfactorily high explained variation to many factors, some of which are hard to interpret economically. In order to overcome this problem we add more structure to the model in such a way that the original approach becomes a special case. This generalized model is based on a Markov chain that switches between different regimes. Each regime is modelled as before by principal components. The regimes represent classes of typical movements like e.g. shifts upwards.

We identify the groups of surface movements by cluster analysis. Cluster analysis aims at constructing homogeneous groups out of heterogeneous large samples. To this end, two fundamental choices about a distance measure and a group building algorithm have to be made. In analogy to the principal components analysis we interpret an observed surface as a sample vector and group the n samples in the rows of a matrix $\mathcal{X}$. As the surfaces are metric variables we choose as distance measure simply the L_2-norm. The L_1-norm gives less weights to outliers while the L_∞-norm only considers maximal outliers. In this way, we get the first distance matrix $d_{ij} = \|x_i - x_j\|_2,\ i, j = 1, \ldots, n$ where x_i denotes the ith row of the matrix $\mathcal{X}$ and corresponds to the ith surface. As cluster algorithm we use a hierarchical method that starts from the finest partition possible and then groups the observations. We apply the agglomerative algorithm that finds the two clusters with the closest distance, puts these two clusters in one cluster and then starts again with a new distance matrix until all clusters are agglomerated into $\mathcal{X}$. We compute the distance between groups by the Ward algorithm: If two groups, P and Q, are united then the distance of this united group to a group R is

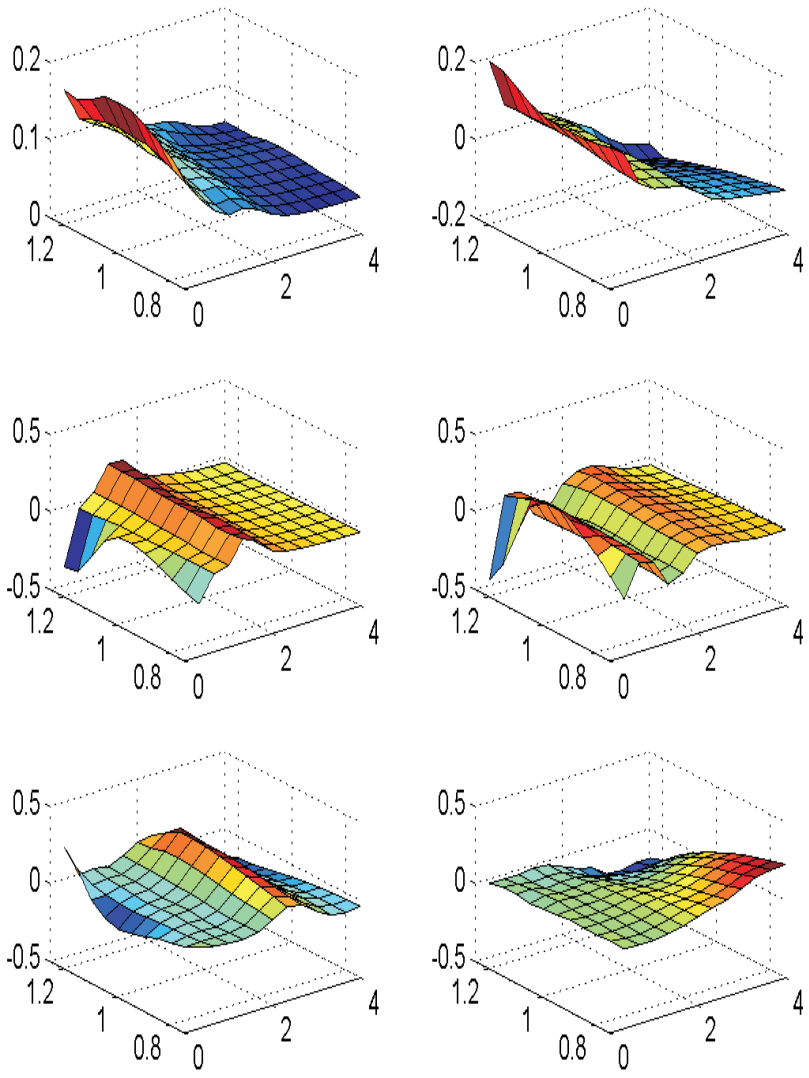

Figure 2.13: Principal components of daily log implied volatility variation.

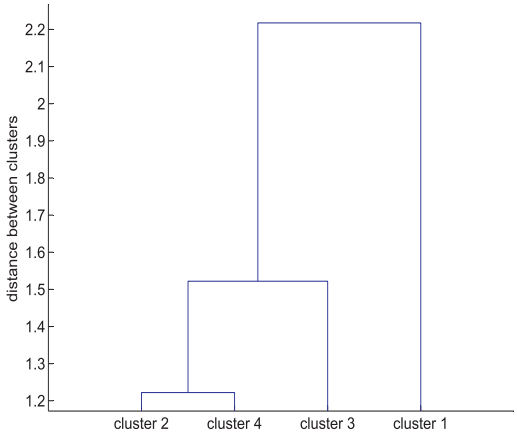

Figure 2.14: Dendrogram of the clusters.

given by

$$d(R, P+Q) = \delta_1 d(R,P) + \delta_2 d(R,Q) - \delta_3 d(P,Q)$$

where $\delta_1 = (n_R + n_P)/n_o,\ \delta_2 = (n_R + n_Q)/n_o,\ \delta_1 = n_R/n_o,\ n_o = n_p + n_Q + n_R,\ n_P = \sum \mathbf{1}(x_i \in P)$ and n_Q and n_R analogue. The Ward method aims at unifying groups in such a way that the variation inside these groups does not increase too much.

We use this cluster analysis to construct groups of similar movements of the implied volatility surface. The resulting grouping is shown in figure 2.14 by a dendrogram which we restrict to four clusters. More clusters are hard to interpret economically. The two fundamental clusters are given by cluster 1 and the cluster consisting of all other observations. This mixed cluster can be divided into cluster 3 on the one hand and cluster 2 and 4 on the other hand. This clustering into three groups is statistically justified because of the big distance between the clusters. We consider another partitioning into four cluster for economic reasons although the distance between cluster 2 and cluster 4 is rather small. The size of each cluster is given in table 2.3. The clusters 1 and 3 are relatively small while most of the observations lie in cluster 2 and 4.

In order to interpret the clusters, we present the mean movement of each group in figure 2.15. The first group represents strong upward movements that also change the term structure by increasing the short level more than the long level. The second cluster contains small upward shifts that change the term structure also a little. The third groups represents in analogy to the first group strong downward movements. But the term structure remains

cluster i	proportion in cluster i
cluster 1	13.3
cluster 2	28.5
cluster 3	08.3
cluster 4	49.9

Table 2.3: Relative sizes of the clusters.

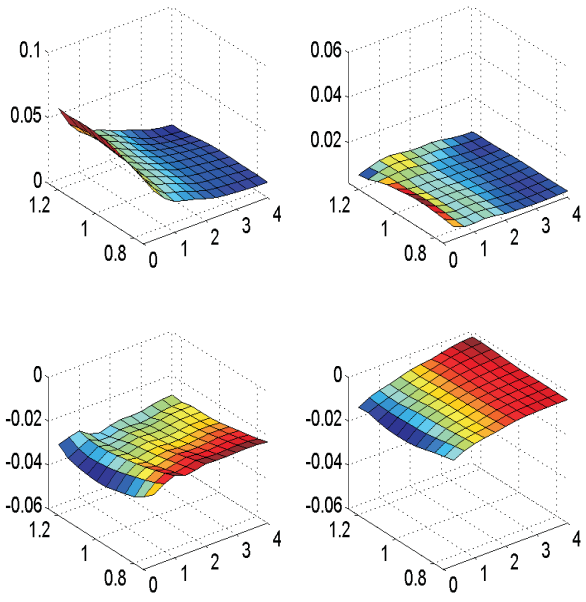

Figure 2.15: Mean surfaces of each cluster (up,left: cluster 1; up,right: cluster 2; down,left: cluster 3; down,right: cluster 4).

relatively constant in this cluster: The short end falls only a bit more than the long end. The last cluster represents small downward movements with some changes in the term structure. The smile effect of implied volatility surfaces which is pronounced only for short times to maturity is vaguely visible at 4 months to maturity. In general, upwards movements are accompanied by a diminution of the smile effect and downward movements lead to a more pronounced smile. The skew is changed stronger by downward movements. The small movements downward seem to make to skew more pronounced while big movements tend to flatten the skew.

In figure 2.16, we show how the clusters are separated by plotting the observed surfaces in the first two principal components that we determined in section 2.3.3. Clusters 1 and 3 lie at the borders and seem different from

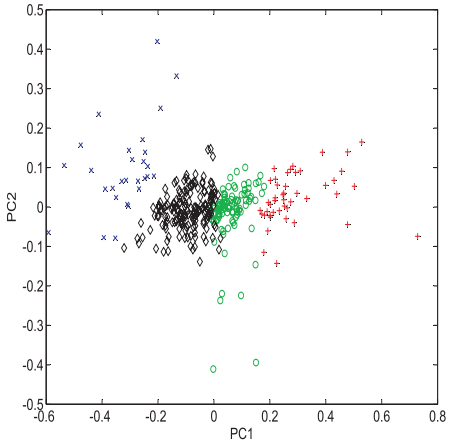

Figure 2.16: Principal components of daily log DAX implied volatility variation, March 2003 - July 2004 (red +: cluster 1, green o: cluster 2, blue x: cluster 3, black d: cluster 4).

principal component	1	2	3
cluster 1	61.9	78.7	85.3
cluster 2	40.1	62.7	76.1
cluster 3	46.9	69.9	79.3
cluster 4	41.6	66.2	79.2

Table 2.4: Cumulative explained variation in the clusters.

the clusters 2 and 4 that are very close to each other. This figure underlines again that clusters 2 and 4 are different not in the statistical sense but in the economic sense representing movements upwards and downwards. Moreover, we see that the clusters are non overlapping.

Each group is modelled by principal components as in section 2.3.3. The first two components of the cluster 1 are presented in figure 2.17 and can be interpreted as level and slope in the term structure dimension. The explained variance of the first three components in each cluster is described in table 2.4. We see that the explained variation in the clusters is in general smaller than in section 2.3.3. But a lot of the overall variation is explained by the clustering.

In order to see how much variation is explained by the clusters and how much variation is explained within the clusters we consider without loss of

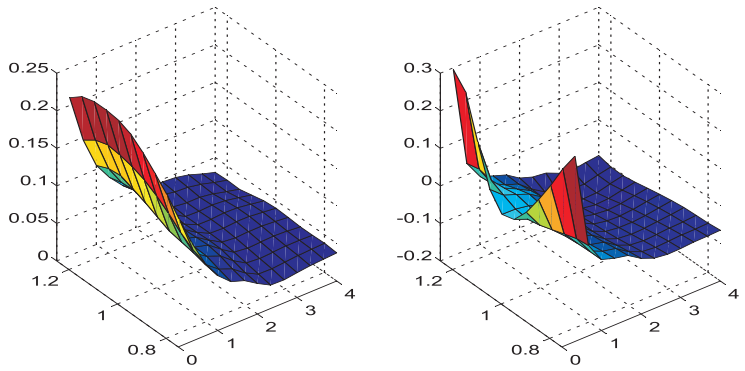

Figure 2.17: First and second principal component in cluster 1.

generality two clusters 1 and 2 where cluster 1 contains the first n_1 observations. Hence cluster 2 contains the second part of all observations $x_1, \ldots, x_n$. For a scalar variable X, we have

$$\begin{aligned}\mathrm{Var}(X) =& \frac{1}{n}\sum_i (x_i - \mu)^2 \\ =& \frac{1}{n}\left\{\sum_{i=1}^{n_1}(x_i - \bar{x}_1 + \bar{x}_1 - \mu)^2 \right. \\ & \left. + \sum_{i=n_1+1}^{n}(x_i - \bar{x}_2 + \bar{x}_2 - \mu)^2\right\}\end{aligned}$$

where μ is the (sample) mean of X and $\bar{x}_1 = n_1^{-1}\sum_{i=1}^{n_1} x_i$ denotes the mean of X in cluster 1 and μ_2 analogue. Because of $\sum_{i=1}^{n_1}(x_i - \bar{x}_1) = 0$ we have

$$\begin{aligned}\mathrm{Var}(X) =& \frac{1}{n}\left\{\sum_{i=1}^{n_1}(x_i - \bar{x}_1)^2 + n_1(\bar{x}_1 - \mu)^2 \right. \\ & \left. + \sum_{i=n_1+1}^{n}(x_i - \bar{x}_2)^2 + (n - n_1)(\bar{x}_2 - \mu)^2\right\} \\ =& \frac{1}{n}\left\{n_1\sigma_1^2 + n_1(\bar{x}_1 - \mu)^2 \right. \\ & \left. + (n - n_1)\sigma_2^2 + (n - n_1)(\bar{x}_2 - \mu)^2\right\}\end{aligned}$$

where $\sigma_1^2 = n_1^{-1}\sum_{i=1}^{n_1}(x_i - \bar{x}_1)^2$ denotes the variance of X in cluster 1 and σ_2^2

number of clusters	2	3	4	5
explained variation	32.9	48.4	58.3	61.8

Table 2.5: Variation explained by clusters.

analogue. Hence, we arrive at

$$\begin{aligned}\mathrm{Var}(X) =& \frac{n_1}{n}\{\sigma_1^2 + (\bar{x}_1 - \mu)^2\} \\ &+ \frac{n - n_1}{n}\{\sigma_2^2 + (\bar{x}_2 - \mu)^2\}\end{aligned}$$

This translates directly to multivariate variables $X = (X_1, \ldots, X_K)$ because of $\mathrm{Var}(X) = \sum \mathrm{Var}(X_k)$:

$$\begin{aligned}\mathrm{Var}(X) =& \frac{n_1}{n}\left(\sigma_1^2 + \|\bar{x}_1 - \mu\|^2\right) \\ &+ \frac{n - n_1}{n}\left(\sigma_2^2 + \|\bar{x}_2 - \mu\|^2\right)\end{aligned}$$

If we explain in each cluster the proportion ρ_i of the variation in the cluster then the explained overall variation can be represented as

$$\frac{n_1}{n}\frac{\sigma_1^2}{\sigma^2}\rho_1 + \frac{n - n_1}{n}\frac{\sigma_2^2}{\sigma^2}\rho_2 + \frac{\|\bar{x}_1 - \mu\|^2 n_1/n + \|\bar{x}_2 - \mu\|^2 (n - n_1)/n}{\sigma^2}$$

where σ^2 denotes the (overall) variance of X. Here, the first sum explains the variation in the clusters and the second sum gives the variation explained by the clustering. An analogue formula holds for more than 2 clusters.

In table 2.5 we show how much of the total variation is explained by the clustering. If we use only two clusters they account for 33% of the variation. Using another cluster increases the explained variance by 15%. A fourth cluster still increases the explained proportion by 10% but a fifth cluster adds only a small explanation of 3%. Hence, we use 4 clusters.

Focusing on 4 clusters the question arises how many principal components should be used in each cluster. In table 2.6 we show how the explained total variation increases with the number of principal components used. For simplification we increase the number of components in all cluster at the same time although insignificant components could be avoided in some clusters probably. One components (for each cluster) explains already 78% of the variation. From table 2.2 we see that the first component in the whole data set explains only 65%. Hence, the clustering increased the explained variance by 13%. Using the first two components (in each cluster) accounts for 87%

number of pc per cluster	1	2	3
explained variation	77.7	87.0	91.7

Table 2.6: Total variation explained by 4 clusters for different numbers of principals components per cluster.

of the overall variance. This corresponds to four principal components for the whole data set. In the same way, three components per cluster explain more than five in the whole data set.

In order to apply the cluster model we have to model the switching between the groups. To this end, we use a (discrete) Markov chain with the groups as state space. Although the transition matrix could change in time we assume a constant transition matrix K, i.e. a time-homogeneous Markov chain. The estimated transition matrix is given by

$$K = \begin{pmatrix} 8.9 & 31.1 & 6.7 & 53.3 \\ 19.0 & 20.0 & 8.4 & 52.6 \\ 17.9 & 35.7 & 17.9 & 28.6 \\ 10.1 & 31.6 & 7.1 & 51.2 \end{pmatrix}$$

and describes the probabilities to change from one group to another group, e.g. the probability to change from group 1 to group 2 is 31.1%. We see that the chain switches most often to groups of small movements, i.e. cluster 2 and 4. Moreover, the chain has the stationary distribution

$$\pi = \begin{pmatrix} 13.1 & 28.5 & 8.3 & 50.0 \end{pmatrix},$$

i.e. $\pi = \pi K$. Hence, the chain forgets where it started and converges to the stationary distribution.

Using this Markov chain model the resulting model for the movements of the implied volatility surfaces is given by

$$\Delta X_t = \sum_{i=1}^{4} \mathbf{1}_{\{M_t=i\}} \left\{ pc_0^i + \sum_{j=1}^{2} \lambda_j^i(t) pc_j^i \right\}$$

where we used two principal components (pc_1^i, pc_2^i) and the mean surface (pc_0^i) for cluster i. The factor loading of the principal component j in cluster i at time t is denoted by $\lambda_j^i(t)$. This model for the movements of implied volatility surfaces gives via equation (2.1) a model for the surfaces themselves. After the estimation of the principal components and the Markov chain the

model is specified by the models for the time series of factor loadings.

As we want to generalize the approach of Cont and da Fonseca (2002) who used AR(1) processes for the time series of factor loadings we also model the factor loadings of each principal component in each cluster by univariate AR(1) processes

$$x_{t+1} = \alpha + \beta x_t + \varepsilon_t.$$

But we observe only one factor loading at each time point because the Markov chain switches between the clusters. Hence, we have to account in the estimation for these missing observations. Given the observation x_t the process at time $t + j + 1$ has the representation

$$x_{t+j+1} = \alpha \frac{\beta^{j+1} - 1}{\beta - 1} + \beta^{j+1} x_t + \sum_{i=0}^{j} \beta^i \varepsilon_{t+i}$$

This can be seen directly by induction. If we do not observe the process at all times $0, \ldots, n$ but only at the times $\theta(i),\ i = 0, \ldots, m$ with an increasing sequence θ then we estimate the parameters α and β of the process by minimizing the least squares distance

$$\sum_{i=1}^{m} \left\{ x_{\theta(i)} - \left(\alpha \frac{\beta^{\theta(i)-\theta(i-1)} - 1}{\beta - 1} + \beta^{\theta(i)-\theta(i-1)} x_{\theta(i-1)} \right) \right\}^2 .$$

This optimization can be done easily by numerical algorithms. By simulation studies we confirmed that AR(1) processes can be estimated in our situation with around $n = 350$ days and our cluster sizes. Moreover, we checked the estimation results for each cluster by estimations from all possible subclusters with one observation less. As we arrived always at similar estimation results this estimation approach seems stable.

2.3.5 Conclusion

Applying a principal components analysis to the movements of implied volatility surfaces with long times to maturity we see that the explanatory power of this approach is limited. In order to make the unexplained variation sufficiently small many components have to be used.

Hence, we propose a generalization of this approach: First, groups of homogeneous movements are constructed by cluster analysis. Then the movements in each group are modelled by a few principal components. Finally, we model the switching between groups by a Markov chain. This approach

has the advantage to require for each group only a few principal components that can be interpreted economically.

The constructed clusters can be described roughly as strong movements upwards, strong movements downwards and small movements. Similar structures have been observed for the returns of stocks. In alternative to our statistical clustering one could define groups economically and then classify the movements by a statistical technique. Future research can analyze which approach leads to more satisfactory economic results.

2.4 Market for Variance Swaps

In the last section, we analyzed the market for European options. This is the most important market because of hedging. But also variance swaps have become an important liquid market for hedging and speculation. Hence, we analyze in this section for these instruments the in- and out-of-sample performance of popular approaches like the Heston model or the Nelson-Siegel parametrization. We observe that the short end of variance swap curves is hard to forecast because of the high variation and curvature in this region. But the random walk is outperformed by models for variance swap prices with long times-to-maturity. Moreover, we conclude that nonparametric methods appear inferior for the forecasting of variance swap curves.

2.4.1 Introduction

In the last 30 years, we have witnessed major advances in the modeling of option prices and great improvements in the estimation of option pricing models. Although there are models like the local volatility model of Dupire (1994) that permit in practice a good and stable fit of observed plain vanilla surfaces these models imply dynamics rarely observed on the markets. Hence, the dynamic evolution of the markets has become a benchmark for models. As modern approaches to option pricing are often based on variance swap markets we analyze in this section the dynamics of variance swap curves with a focus on factor modeling.

Variance swap markets have become quite liquid and these products serve – just like the plain vanilla options – as hedging instruments for some modern options like calls on realized variance. Thus, the analysis of variance curves is as important as the study of the evolution of implied volatility surfaces. As variance swap prices can be derived from implied volatility surfaces by the results of Neuberger (1992) variance swap curves reflect basically the term structure of implied volatility surfaces. Thus, understanding the evolution of

variance curves helps also modeling the plain vanilla market.

In this section, we analyze modeling approaches for variance swap curves focusing on the forecasting perspective. First, we consider the model of Heston (1993) because it can be regarded as a benchmark approach for the popular stochastic volatility models in option pricing. When we fix the mean reversion speed it leads to a two factor structure for variance swap curves. Moreover, we consider a generalization that gives a reparametrization of the model of Nelson and Siegel (1987). In addition, we fit a semiparametric model in order to check the structure of the parametric models. The in-sample fit of all models appears unbiased so that the prices are correctly replicated in the mean. Moreover, the variation of the in-sample errors (measured in volatility) is acceptably small because the inter-quartile-ranges are smaller than 20 basis points. The Nelson-Siegel approach is better than the Heston model for long times-to-maturity and the semiparametric model seems to be slightly better than the parametric approaches. Out-of-sample, the random walk outperforms the models for short time forecasts because of its perfect in-sample fit. But for longer forecasts the parametric models outperform the random walk for long times-to-maturity. Moreover, the semiparametric approach has an unsatisfactory out-of-sample fit and the Heston model without parameter forecasts seems also inferior to the dynamic Heston or Nelson-Siegel model.

We proceed as follows. In section 2.4.2, we give a description and short derivation of the modeling framework, which comprises the Heston model, the Nelson-Siegel parametrization and a semiparametric approach. In section 2.4.3, we conduct an empirical analysis, describing the data, estimating the models and forecasting the variance curves. In addition to the models, we consider some other natural approaches for forecasting. In section 2.4.4, we conclude.

2.4.2 Modeling the Term Structure

In this section, we introduce variance swaps and explain the construction of variance swap curves. Moreover we discuss the stochastic volatility model of Heston (1993) and derive its approximation of variance swap prices. Besides this two parameter approach, we consider a generalization to three parameters analog to the yield curve model of Nelson and Siegel (1987). In addition, we use a semiparametric factor model in order to analyze the parametric assumptions.

Variance swap curves

Variance swaps are forward contracts on future realized variance. They exchange at expiration the realized annualized variance of the log returns of an underlying against a predefined strike. The contracts we consider are based on a zero mean of the returns and $c = 252$ trading days for annualization with daily sampling. Hence, given an underlying (S_t), the payoff of a variance swap (with zero strike) for the period $[0, T]$ with business days $0 = t_0 < \ldots < t_n = T$ is given by

$$\sigma_R^2(T) \stackrel{\text{def}}{=} \frac{c}{n} \sum_{i=1}^{n} (\log S_{t_i} - \log S_{t_{i-1}})^2. \tag{2.2}$$

If we denote the price of (not annualized) variance by $V(T)$ then the price of a variance swap is given by $V(T)/T$. These prices are often quoted in volatility strikes, i.e. $\sqrt{V(T)/T}$, which shows how closely variance swaps are related to volatility swaps that have payoff profiles $\sqrt{\sigma_R^2}$. Actually, there is a variety of options that depend on realized variance, e.g. calls on realized variance with payoffs $\max\{\sigma_R^2 - K, 0\}$. Hence, variance swaps are not only used for directly speculating on realized variance but they serve also as hedging instruments for more difficult products.

At a point in time we observe N prices of variance swaps $V(x_i)/x_i$ with expiries $x_1, \ldots, x_N$. The variance swap curve at this time is given by the mapping $T \mapsto V(T)/T$. We call V the variance curve and V' the forward variance curve. Many approaches for modeling variance swap curves are based on forward variance curves, e.g. Bergomi (2005). But in practice, variance curves or forward variance curves are not observed. Instead, they must be estimated from a discrete set of observed variance swap prices.

Besides the prices of variance swaps observed on markets, theoretical prices can be derived from implied volatility surfaces in a way that is quite model independent:

$$V(T)/T = \frac{2}{T} \left\{ 1 + rT - e^{rT} + e^{rT} \int_0^{S_0} \frac{P(K)}{K^2} dK + e^{rT} \int_{S_0}^{\infty} \frac{C(K)}{K^2} dK \right\} \tag{2.3}$$

where S_0 denotes the spot and $P(K)$ $[C(K)]$ a put [call] price for a strike K, see e.g. Demeterfi et al. (1999). In this way, additional prices of variance swaps can be derived.

Nevertheless, we have only a finite number of prices of the variance swap curve. As the times-to-maturity of the implied volatility surface change every

day the observations lie moreover on a moving grid. For comparing the in- and out-of-sample performance of models on different days we need the variance swap curves on a fixed grid. Local polynomial regression like the Nadaraya-Watson estimator tends to oversmooth the data because the few observed prices require a relatively high bandwidth. This underfitting can be avoided by smoothing splines which allow also smaller bandwidths. But their boundary behaviour is hard to control. Hence, we use cubic splines with constraints on the first derivatives at the end points of the data. Using market prices of variance swaps together with additional synthetic prices we construct the variance curves by cubic splines. In this way, also the other curves, e.g. the forward variance curve, are determined. We interpolate the variance swap curves (not quoted in volatility) because we estimate the models from these data, see section 2.4.3.

Modeling variance swap curves

In this part, we describe the Heston model and its price formula for variance swaps. Then we generalize this in analogy to Nelson-Siegel's yield curve model. Finally, we consider a semiparametric model for variance swap curves.

The stochastic volatility model of Heston is given by

$$\begin{aligned} \frac{dS_t}{S_t} &= r dt + \sqrt{\zeta_t} dW_t^{(1)} \\ d\zeta_t &= \kappa(\theta - \zeta_t)dt + \nu\sqrt{\zeta_t} dW_t^{(2)} \end{aligned}$$

where $W^{(1)}$ and $W^{(2)}$ are correlated Wiener processes. The parameter θ is the long variance level because the variance process ζ fluctuates around it. The parameter κ controls how strong the variance process is pulled back to this long variance level. Hence, κ is called mean reversion speed. The volatility of variance is given by ν. Additional parameters are the initial value of the variance process ζ_0 and the correlation between the Wiener processes.

Approximating the payoff (2.2) by the quadratic variation of the logarithm of the underlying we can derive the prices of variance swaps as

$$\begin{aligned} V(T)/T &= E\left(\frac{c}{n}\sum_{i=1}^{n}(\log S_{t_i} - \log S_{t_{i-1}})^2\right) \\ &= \frac{1}{T}E\left(\int_0^T \zeta_t dt\right) \\ &= \theta + (\zeta_0 - \theta)\frac{1-\exp(-\kappa T)}{\kappa T}. \end{aligned} \tag{2.4}$$

Hence, only the initial variance ζ_0, the long variance θ and the mean reversion speed κ determine the variance swap prices. The other two parameter of the Heston model, the volatility of variance and the correlation, control the smile of the implied volatility surfaces. Because of the representation (2.3) these two parameters do not enter the formula of variance swap price.

The corresponding model for the forward variance curve is given by

$$V'(T) = \theta + (\zeta_0 - \theta)\exp(-\kappa T). \tag{2.5}$$

This forward variance curve model implies exactly the above variance swap prices because of the constraint $V(0) = 0$. This model for the forward variance curves is also called linearly mean-reverting (forward) variance curve model, see Bühler (2006).

For a fixed mean reversion speed κ this model has two parameters z_1 and z_2 that we get from the parametrization $z_1 = \theta$ and $z_2 = \zeta_0 - \theta$. Using this parametrization we recognize (2.5) as a short form of the Nelson-Siegel parametrization for forward rates. Hence, we consider also the full Nelson-Siegel parametrization for the forward variance curve

$$V'(T) = z_1 + z_2\exp(-\kappa T) + z_3\kappa T\exp(-\kappa T)$$

This model is also called the double mean-reverting (forward) variance curve model and the variance swap prices $V(T)/T$ are given in it by

$$V(T)/T = z_1 + z_2\frac{1-\exp(-\kappa T)}{\kappa T} + z_3\left\{\frac{1-\exp(-\kappa T)}{\kappa T} - \exp(-\kappa T)\right\} \tag{2.6}$$

because of

$$\frac{V(T)}{T} = \frac{1}{T}\int_0^T V'(t)dt.$$

While the linearly mean-reverting model is basically the Heston model, the second approach was considered by Bühler (2006) who analyzed conditions for an arbitrage free joint market of variance swaps and stock. His considerations imply that the mean reversion speed κ should be constant. In practice, a constant mean reversion speed is important for stability of the parameters. Because of these theoretical and practical reasons we fix this parameter and use $\kappa = 2$ as in Bergomi (2004). This fixing makes the models linear in the parameters and hence simplifies the estimation considerably.

In addition to these two parametric models, we consider a semiparametric approach in order to analyze how severe the constraints of the functional

forms are. For this purpose, we apply a semiparametric approach described in Fengler (2005). This method approximates variance swap curves by unknown basis functions that have to be estimated from the data. These nonparametric factors offer more flexibility than the factors in the Heston or the Nelson-Siegel model. The factor loadings correspond to the parameters in those models.

The semiparametric model regresses the logarithm of variance swap prices on time-to-maturity. Let $Y_{t,j}$ be the logarithm of the j-th observed variance swap price on day t with maturity $T_{t,j}$. Then the model regresses $Y_{t,j}$ on $T_{t,j}$ by

$$Y_{t,j} = m_0(T_{t,j}) + \sum_{l=1}^{L} z_{t,l} m_l(T_{t,j}) + \varepsilon_{t,j},$$

where m_0 is an invariant basis function, m_l $(l = 1, ..., L)$ are the basis functions and $z_{t,l}$ are the factor weights depending on time t. As the parametric models have two and three parameters we use also $L = 3$ basis functions in the semiparametric model. We describe the estimation procedure in section 2.4.3 where we use the data.

The parameters z are estimated on each day so that we get time series of estimated parameters (z_t). The parameters (z_{1t}, z_{2t}) in the Heston model, the parameters (z_{1t}, z_{2t}, z_{3t}) in the Nelson-Siegel framework and the weights (z_{1t}, z_{2t}, z_{3t}) in the semiparametric approach can all be interpreted as latent dynamic factor loadings for variances swap curves. As the Heston price formula (2.4) is a special case of the Nelson-Siegel formula (2.6) we discuss only the factors in the Nelson-Siegel approach and the semiparametric model.

We start with the Nelson-Siegel parametrization. The factor on z_{1t} is the constant 1. As this factor does not decay to zero in the long run it can be interpreted as a long-term factor. The factor on z_{2t} is $\{1-\exp(-\kappa T)\}/(\kappa T)$. This function is monotonically decreasing from 1 to 0. As it influences mainly the short end of the curve it can be interpreted as a short-term factor. Besides these two factors the Nelson-Siegel model controls also the medium-term. The factor on z_{3t} is $\{1-\exp(-\kappa T)\}/(\kappa T) - \exp(-\kappa T)$. This mapping increases monotonically from 0 to a peak and then decreases to zero in the long-term in a similar way as the second factor. This form explains the interpretation as a medium-term factor. These three factors are presented in figure 2.18.

Moreover, the parameters have interpretations as level, slope and curvature of the curves: As an increase in z_1 increases the whole curve by the same amount the factor on z_1 represents the level of the curve. An increase

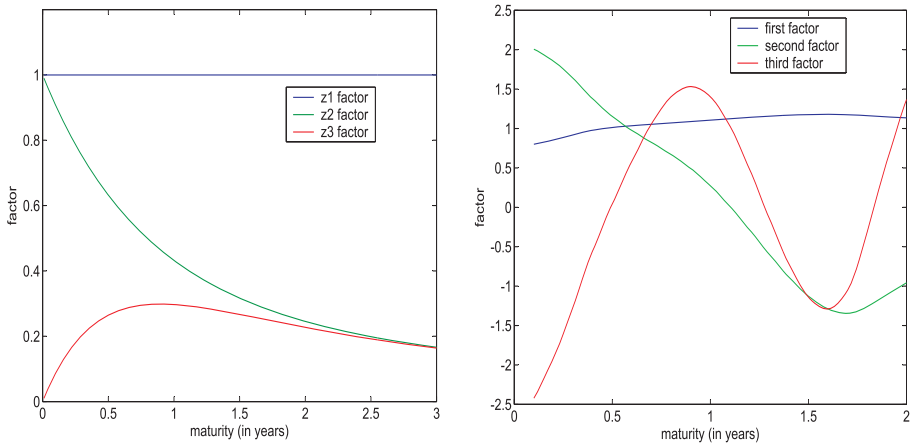

Figure 2.18: Factors for variance curves in the Nelson-Siegel model (left) and in the semiparametric model (right).

of the short-term factor increases the curve more at the short end than at the long end. Hence it controls the slope of the curve. Finally, the third factor moves the middle of the curve while keeping the ends (almost) fixed. In this way it changes the curvature of the curve. Hence, the difference between the Heston and the Nelson-Siegel model is the capability to control the curvature directly.

The factors of the semiparametric model are shown in figure 2.18. They can be interpreted again as level, slope and curvature. In contrast to the factors of the parametric models, they are sometimes negative because we regress the logarithm of the prices. (The exponential transformation ensures positivity of the prices.) The third factor does not show the typical form of curvature because the concave form changes for long maturities into a convex form. This special shape means that a stronger curvature of the curves is often accompanied by a higher price for 2 years to maturity.

2.4.3 Forecasting the Term Structure

In this section, we describe the data, estimate the factor loadings, model and forecast them. Finally, we compare the goodness of the forecasts of the variance swap curves.

The data

The data set studied contains prices of variance swaps on the S&P 500 index between 1 October 2003 and 30 November 2005. These swaps use daily closing prices of the index, have 252 business days as annualization factor and assume a zero mean for the calculation of the variance of the returns. Each day we observe the mid market prices of variance swaps with 1 year and 2 years to maturity. These data represent real trading prices of a large financial institution. In addition, we have synthetic prices of variance swaps derived from implied volatility surfaces via (2.3). In this way, we have in the mean around 7 observations each day. The maturities always cover the range from 0.25 to 2 years and we have no observations with more than 2 years to maturity.

The parameters of the Heston and the Nelson-Siegel model are estimated from these original data. The semiparametric model requires more observations per day for a stable calibration of the nonparametric factors so that we have to interpolate or smooth the data. Moreover, we want to compare the data on a fixed grid in the analysis of the in- and out-of-sample performance. As discussed in section 2.4.2 we use cubic splines with conditions on the derivatives at the end points. In this way, we can construct the whole curves although we use only the maturities 0.25, 0.5, 1.0 and 2.0 years for the performance analysis.

In figure 2.19, we present the variance swap curves from the interpolation with cubic splines and the variance swap curves quoted in volatility strikes. We estimate and forecast the variance swap curves because they are easier to estimate then the curves in volatility strikes. But in industry the prices are normally considered in volatility strikes because of the intuition. Hence, we analyze the results in volatility strikes. The figures show at the beginning a market regime of high expected realized volatility and afterwards lower prices. Moreover, the curves fluctuate stronger for short times to maturity. In figure 2.20, we show the corresponding variance and forward variance curves. The variance curves show clearly the variation in the level of the variance swap curves. But also the curvature is visible although it is more apparent in the forward variance curves.

We provide some descriptive statistics of the variance swap curves in volatility strikes in table 2.7. The mean curve that is also shown in figure 2.21 is increasing and concave. In some markets we can observe sometimes an inverse curve that is falling at the beginning. Moreover, the table demonstrates that the short end of the curves has a higher variation – in absolute as well as in relative terms. Hence, the curves are harder to model in- and especially out-of-sample in this region. The forecasting difficulties are also

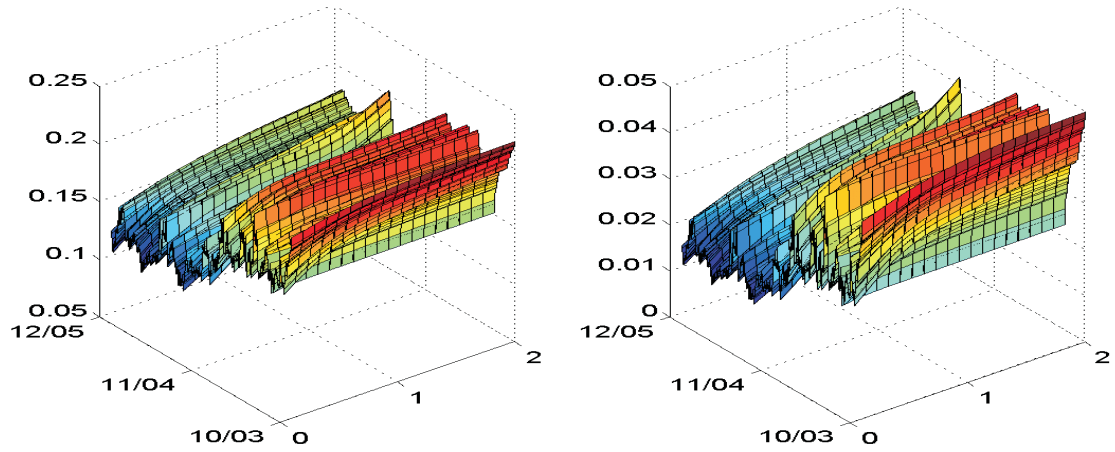

Figure 2.19: Daily variance swap curves quoted in volatility strikes (left) and variance swap curves (right), 01/10/03 - 30/11/05. (left axis: observation day, right axis: time to maturity)

Maturity	Mean	Std.dev.	Min.	Max.	$\hat{\rho}(1)$	$\hat{\rho}(5)$	$\hat{\rho}(20)$
0.25	17.1	1.62	13.2	21.2	94.2	76.0	34.3
0.50	18.0	1.54	14.3	21.5	96.1	83.0	41.6
1.00	18.7	1.50	15.0	22.0	97.6	87.4	46.3
2.00	19.2	1.48	15.1	22.3	98.2	90.0	46.8

Table 2.7: Descriptive statistics of interpolated variance swap curves in volatility, 10/01/03 - 30/11/04.

shown by the autocorrelations that decay faster for small times to maturity.

Fitting the variance swap curves

The modeling of variance swap curves is often based on the forward variance curves. But these curves cannot be observed directly and hence have to be constructed from the finite number of variance swap prices. Because of the dependence on this construction forward variance curves are not useful for the estimation of models. Variance curves start from zero and increase monotonically. Estimating models from variance curve data gives in general an unsatisfactory fit for short maturities because the absolute error is always rather small for short maturities. Calibrating models to variance swap curves in volatility strikes minimizes the intuitive error in volatility points. But this estimation requires numerical methods because of the square root transformation. The estimation based on variance swap curves minimizes

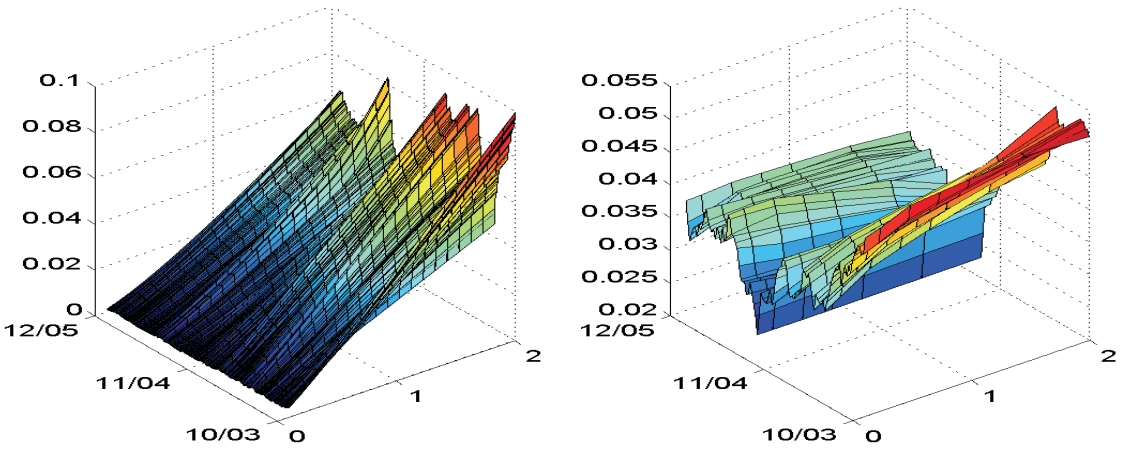

Figure 2.20: Daily variance curves (left) and forward variance curves (right), 01/10/03 - 30/11/05. (left axis: observation day, right axis: time to maturity)

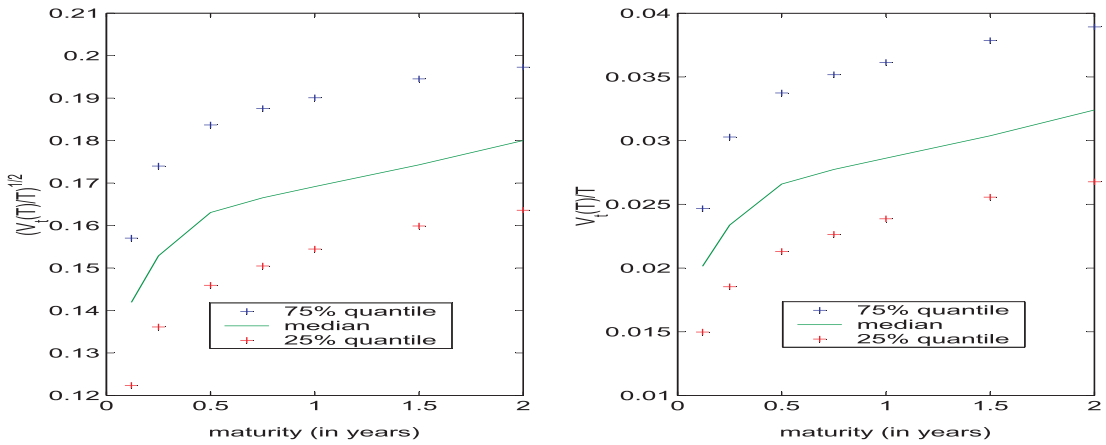

Figure 2.21: Median data-based variance swap curves (left) quoted in volatility strikes (right) with pointwise interquartile range.

the real price differences. Moreover, this estimation can be carried out easily by the (weighted) least squares method. Hence, we calibrate the models to the variance swap prices.

These prices are given in the Heston model by

$$z_1 + z_2 \frac{1 - \exp(-\kappa T)}{\kappa T}$$

and in the Nelson-Siegel model by

$$z_1 + z_2 \frac{1 - \exp(-\kappa T)}{\kappa T} + z_3 \left\{ \frac{1 - \exp(-\kappa T)}{\kappa T} - \exp(-\kappa T) \right\}.$$

In the Nelson-Siegel approach for interest rates it is common to fix the parameter κ. Hence, we also fix this parameter κ in our Nelson-Siegel parametrization for variance swaps. Moreover, it is practice to fix this parameter for stability reasons in the Heston model for the modeling, pricing and hedging of options. Thus, we use $\kappa = 2$ as in Bergomi (2004). Keeping this parameter constant allows us to estimate the parameters by ordinary least squares. But this method does not take into account parameter constraints like the positivity of the initial variance. Therefore we checked after the estimation that these constraints were indeed fulfilled. In order to take into account the distribution of the observation in the time-to-maturity dimension we use weighted least squares. In this way, we estimate each day the parameters of the Heston and the Nelson-Siegel model and get for each model a time series $(\hat{z}_t)$ of parameters.

We estimate the nonparametric factors in the semiparametric model from a training set that we choose as the first 14 months of our time series. On day t we use the logarithm $Y_{t,j}$ of the observed variance swap price with maturities $T_{t,j}$. The factors or basis functions $\widehat{m}_l$ and the factor loadings $\widehat{z}_{t,l}$ are then estimated by minimizing the following least squares criterion $(\widehat{z}_{t,0} = 1)$:

$$\sum_{t=1}^{n} \sum_{j=1}^{J_t} \int \left\{ Y_{t,j} - \sum_{l=0}^{L} \widehat{z}_{t,l} \widehat{m}_l(u) \right\}^2 K_h(u - T_{t,j})\, du,$$

where K_h denotes a kernel function. The minimization procedure searches through all functions $\widehat{m}_l : \mathbb{R} \to \mathbb{R}$ and time series $\widehat{z}_{t,l} \in \mathbb{R}$ by an iterative procedure. Afterwards the estimates are orthogonalized and normalized, see Fengler (2005) for details. The estimated factors are plotted in figure 2.18. For this estimation we used a bandwidth that is optimal in a sense of the Akaike information criterion, see Fengler (2005) for details. The factors can

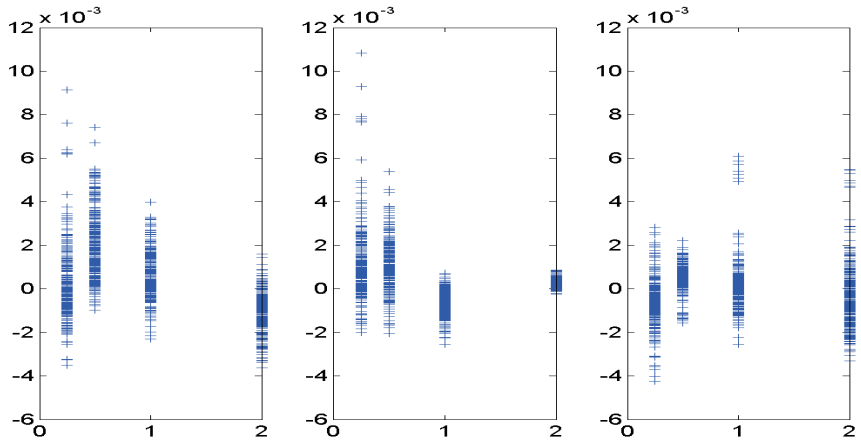

Figure 2.22: Residuals of variance swap curves in volatility, 01/10/03 - 30/11/05. (left: Heston, middle: Nelson-Siegel, right: semiparametric model)

be interpreted as in the parametric models as level, slope and curvature, see section 2.4.2.

Information about the in-sample fit of the models is presented in figure 2.22. It shows that the Heston model is rather unbiased because the residuals are centered around zero. Moreover, the variation of these errors is rather constant over the different maturities. The fit seems acceptable although not perfect for the liquid market of variance swaps. Moreover, the means of the residuals seem to show that the Heston model has problems in replicating the concavity of the curves. The Nelson-Siegel model is a generalization of the Heston model for variance swap prices. Hence, its in-sample fit should be better. Our estimation was based on the real variance swap prices. In this scale, the Nelson-Siegel model leads to an improvement of about 30%. Transformed to volatility strikes this improvement is about 20%. Figure 2.22 documents that mostly the long times to maturity are better fitted because of the smaller variation. Especially the prices for 2 years to maturity are unbiased with a quite small variation. The residuals in the semiparametric model are also shown in figure 2.22. They are unbiased as in the parametric models and their variation is slightly smaller than for the parametric models for short times-to-maturity.

Table 2.8 presents some descriptive statistics of the residuals. The statistics confirm the interpretation of the figure 2.22. The mean and median are (except for one case) below 10 basis points. The standard deviation and the quartiles demonstrate how well the Nelson-Siegel model fits the prices

Maturity	Mean	Std.dev.	Median	25% q.	75% q.	Min.	Max.
Heston							
0.25	3	17	0	-8	11	-35	91
0.50	21	15	20	10	30	-10	74
1.00	6	10	5	0	12	-23	40
2.00	-10	9	-10	-16	-5	-36	16
Nelson-Siegel							
0.25	11	18	8	0	20	-20	108
0.50	11	12	11	3	19	-20	54
1.00	-6	6	-6	-10	-2	-26	7
2.00	2	2	2	1	4	-2	9
semiparam.							
0.25	7	12	-8	-13	-1	-42	28
0.50	5	6	-5	2	9	-16	22
1.00	1	12	1	-5	4	-26	61
2.00	-4	15	-7	-15	3	-33	55

Table 2.8: Descriptive statistics of the residuals of variance swap curves in volatility (in basis points).

with long times to maturity. We can conclude that the Nelson-Siegel model improves the Heston model for long times to maturity.

In figure 2.23 we show the time series of the estimated factor loadings in the Heston model, in the Nelson-Siegel model and in the semiparametric factor model. The time series of the level is always positive and decreases slightly over time. This reflects the lower variance swap prices at the end of our period, see figure 2.19. The high correlation of corresponding factors between the models is also apparent. Moreover, the level loadings of the three models are of the same magnitude because the factors also similar, see figure 2.18. On the other hand, the second and third factors are of different magnitude. Hence, we have rescaled the loadings to make them comparable. These factors show again a high correlation. This underlines that the factors have a common interpretation as level, slope and curvature. The figure demonstrates also that the Heston and the Nelson-Siegel model lead to quite similar loadings. In addition to these estimated loadings, one can define empirical loadings. For example, the level can be identified with the highest variance swap price, i.e. the variance swap price for 2 years to maturity. Such empirical factors are quite similar to the estimates, see e.g.

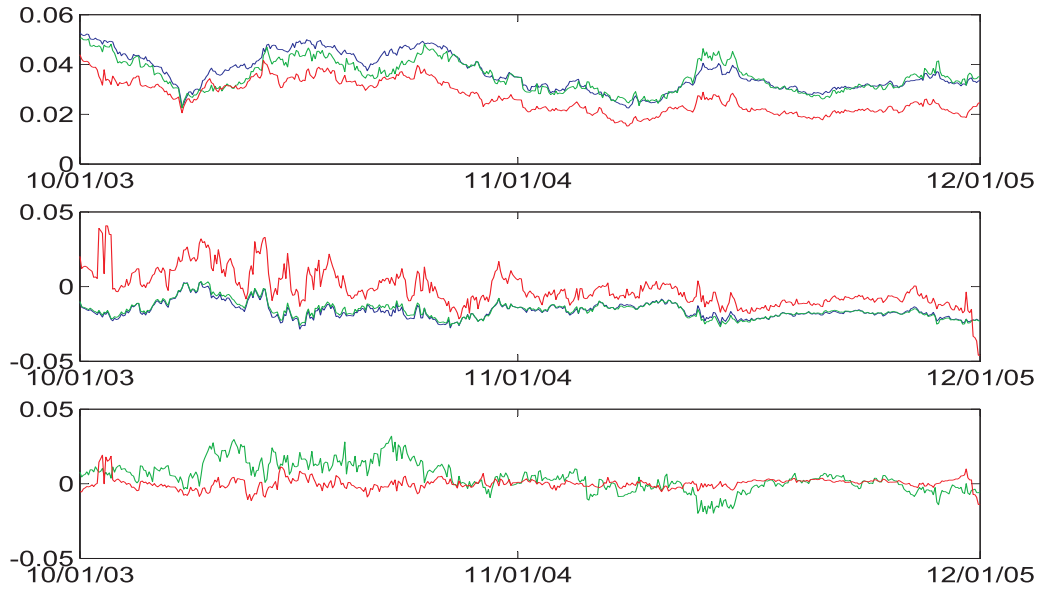

Figure 2.23: Time series of first (upper), second (middle) and third factor loading (lower), 10/01/03 - 12/01/05. (blue: Heston, green: Nelson-Siegel, red: semiparametric model)

Diebold and Li (2006). Descriptive statistics on these factor loading series are presented in table 2.9 that also contains information on the autocorrelations. Augmented Dickey-Fuller tests did not detect unit roots in the time series except for the slope in the semiparametric model.

Modeling and forecasting the factor loadings

Autoregressive processes can be regarded as a simple standard approach for time series modelling because of their good results relative to the parsimonious modeling. They were applied in a variety of contexts and also in finance. Diebold and Li (2006) analyzed the forecasting of yield curves by factor models and used univariate AR(1) processes for modelling the time series of factor loadings. Cont and da Fonseca (2002) considered the dynamics of implied volatility surfaces in a principal components analysis and modelled the resulting time series of factor loadings also by univariate AR(1) processes. Hence, we also use univariate AR(1) processes to model the time series of factor loadings $(\hat{z}_{i,t})$ for $i = 1, 2, 3$.

The resulting forecasts of the variance swap curves τ days ahead are then

Factor	Mean	Std.dev.	Median	$\hat{\rho}(1)$	$\hat{\rho}(5)$	$\hat{\rho}(20)$
Heston						
z_1	367	73	348	99.2	95.1	73.0
z_2	-162	54	-171	93.0	75.5	38.0
Nelson-Siegel						
z_1	353	63	346	97.4	91.0	54.1
z_2	-158	57	-164	94.6	80.0	44.0
z_3	44	96	32	91.8	82.3	58.7
semiparam. model						
z_1	267	65	256	96.7	92.0	78.1
$10z_2$	-14	120	-43	85.5	66.0	41.1
$100z_3$	12	381	17	67.3	33.9	-9.1

Table 2.9: Descriptive statistics of the factor loadings.

given on day t by

$$\widehat{V_{t+\tau}(T)}/T = \hat{z}_{1,t/t+\tau} f_1(T) + \hat{z}_{2,t/t+\tau} f_2(T) + \hat{z}_{3,t/t+\tau} f_3(T)$$

where $\hat{z}_{i,t/t+\tau}$ are the forecasts of the i-th factor loading τ days ahead on day t and f_1, f_2, f_3 are the factors.

We use the factor loadings of the first 14 months of our data set to estimate the univariate AR models. The remaing 12 months of the data are used for validation. In figure 2.24, we show the autocorrelation functions of the residuals of the AR(1) models of the factor loadings. As only few autocorrelations lie outside the 95% confidence intervals we conclude that the modeling is acceptable for the estimated factor loadings. Only the level time series in the Nelson-Siegel shows too many high autocorrelations. But we do not consider a higher order model because these autocorrelations exceed the confidence intervals only slightly.

In tables 2.10 - 2.12, we show the results for 1 month, 2 months and 4 months ahead forecasts. As we have only one year for validation we do not consider longer periods. In addition to the three models considered so far we analyze two simple "competitors", the static Heston model and the random walk.

The two benchmark models are:

- The static Heston model
  In industry, the Heston model is usually applied without forecasting the parameters. On every day the model is calibrated to observed

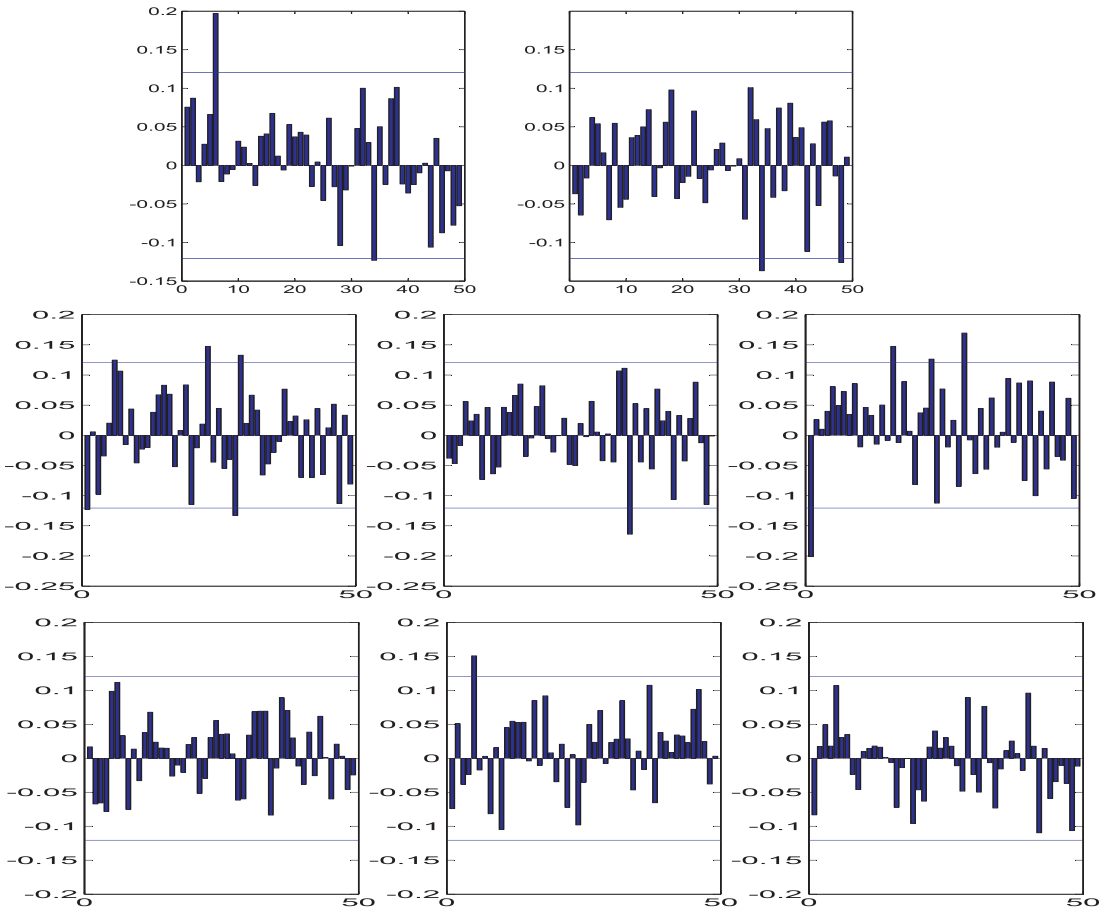

Figure 2.24: Factor loadings in the Heston model (upper), in the Nelson-Siegel model (middle) and in the semiparametric factor model (lower). (left to right: level, slope, curvature)

plain vanilla prices and other prices are calculated on the basis of these fixed parameters. The forecasts τ weeks ahead in this model are

$$\widehat{V_{t+\tau}(T)}/T = \frac{V_t(T+\tau) - V_t(\tau)}{T}$$

where V_t denotes the variance curve at time t.

- The random walk
  This natural benchmark model proposes that the variance swap curves do not change:

$$\widehat{V_{t+\tau}(T)}/T = V_t(T)/T$$

The forecast errors at time $t+\tau$ are defined as the difference between the observed variance swap curve and the forecasted curve (in volatility):

$$\sqrt{V_{t+\tau}(T)/T} - \sqrt{\widehat{V_{t+\tau}(T)}/T}$$

for $T = 0.25, 0.5, 1.0$ or 2.0 years. In tables 2.10 - 2.12 we present some descriptive statistics of these errors. Figure 2.25 shows the errors of the random walk model, i.e. the variation in time of the real curves. As these benchmark errors do not seem to follow a normal distribution we focus in the analysis mainly on robust statistics like the median.

Short time forecasts (less than 1 month ahead) are not reported because for such forecasts the random walk model is in general the best approach. The 1-month ahead forecast errors are described in table 2.10. All models are negatively biased as is shown by the median. This can be interpreted as a general decrease in the price level. The smallest bias has the random walk whose errors also have the smallest variation. This variation, that can be measured by the interquartile range, is roughly the same for the other models. The semiparametric model can be regarded as the best of these models because of its small bias for $0.5, 1.0$ and 2.0 years to maturity.

Out-of-sample results for 2-months ahead are presented in table 2.11. The bias of the whole curve is again smallest for the random walk. Moreover, the bias of the random walk is positive while all other models overestimate the curve and have a negative bias. The bias of the random walk is worst for 2 years to maturity. The other models have in contrast the smallest bias for this time to maturity. The bias of the semiparametric approach is clearly the worst and the Heston and Nelson-Siegel models outperform the static Heston model in terms of the bias. The variation of the errors is similar for all models because the interquartile ranges all vary between 2% and 2.5%.

Maturity	Mean	Std.dev.	Median	25% q	75% q	Min.	Max.
Heston							
0.25	-17	179	-34	-157	93	-363	447
0.50	-13	157	-24	-130	86	-329	378
1.00	-28	148	-31	-132	91	-324	337
2.00	0	148	3	-106	85	-306	400
Nelson-Siegel							
0.25	-48	159	-46	-167	69	-386	317
0.50	-39	143	-35	-134	54	-366	258
1.00	-48	139	-18	-147	32	-382	259
2.00	-16	140	-7	-126	56	-331	329
static Heston							
0.25	-53	140	-45	-157	58	-375	247
0.50	-37	124	-32	-128	61	-325	213
1.00	-39	117	-44	-125	56	-283	204
2.00	-1	121	-11	-94	82	-257	321
semiparam. model							
0.25	-65	148	-67	-183	71	-347	162
0.50	-10	134	-8	-102	67	-307	173
1.00	-17	132	-17	-106	69	-307	140
2.00	10	131	2	-67	74	-288	201
random walk							
0.25	-13	136	-5	-102	91	-345	278
0.50	-10	120	-9	-90	87	-290	243
1.00	-4	114	-10	-86	88	-273	226
2.00	1	123	-2	-88	83	-243	323

Table 2.10: Out-of-sample 1-month-ahead forecasting results of variance swap curves in volatility strikes (in bp).

Maturity	Mean	Std.dev.	Median	25% q	75% q	Min.	Max.
Heston							
0.25	-67	231	-51	-218	41	-518	508
0.50	-56	221	-37	-187	55	-487	476
1.00	-64	218	-35	-194	45	-495	398
2.00	-30	227	-7	-159	83	-450	445
Nelson-Siegel							
0.25	-79	189	-65	-237	-33	-537	239
0.50	-71	188	-46	-218	-11	-551	225
1.00	-70	191	-43	-220	-2	-561	177
2.00	-49	193	-32	-170	52	-488	250
static Heston							
0.25	-89	170	-81	-234	31	-459	420
0.50	-61	157	-55	-197	46	-402	390
1.00	-49	161	-37	-168	47	-386	376
2.00	2	177	2	-147	82	-328	458
semiparam. model							
0.25	-173	202	-161	-308	-70	-679	374
0.50	-118	200	-97	-223	-18	-615	369
1.00	-129	210	-106	-238	-24	-618	357
2.00	-102	230	-92	-222	-4	-602	452
random walk							
0.25	-6	170	14	-147	126	-378	490
0.50	-2	153	5	-133	98	-328	450
1.00	7	157	16	-103	103	-301	436
2.00	16	179	19	-127	105	-335	464

Table 2.11: Out-of-sample 2-months-ahead forecasting results of variance swap curves in volatility strikes (in bp).

Maturity	Mean	Std.dev.	Median	25% q	75% q	Min.	Max.
Heston							
0.25	-96	195	-67	-216	12	-520	415
0.50	-75	187	-50	-202	35	-459	389
1.00	-74	174	-46	-193	33	-445	332
2.00	-30	151	-17	-121	59	-362	337
Nelson-Siegel							
0.25	-119	111	-92	-196	-40	-406	98
0.50	-92	113	-66	-163	-6	-373	101
1.00	-87	108	-55	-150	-2	-368	89
2.00	-42	79	-33	-83	15	-260	107
static Heston							
0.25	-150	130	-142	-233	-56	-431	141
0.50	-96	124	-90	-179	-7	-375	172
1.00	-58	127	-30	-150	30	-352	177
2.00	19	143	52	-99	117	-299	293
semiparam. model							
0.25	-218	188	-180	-336	-87	-648	160
0.50	-155	197	-116	-280	-29	-617	221
1.00	-163	200	-119	-278	-51	-623	213
2.00	-127	193	-83	-238	-26	-571	255
random walk							
0.25	4	144	16	-78	106	-324	293
0.50	17	121	24	-59	102	-248	298
1.00	35	123	68	-36	124	-260	264
2.00	51	152	96	-69	160	-309	323

Table 2.12: Out-of-sample 4-months-ahead forecasting results of variance swap curves in volatility strikes (in bp).

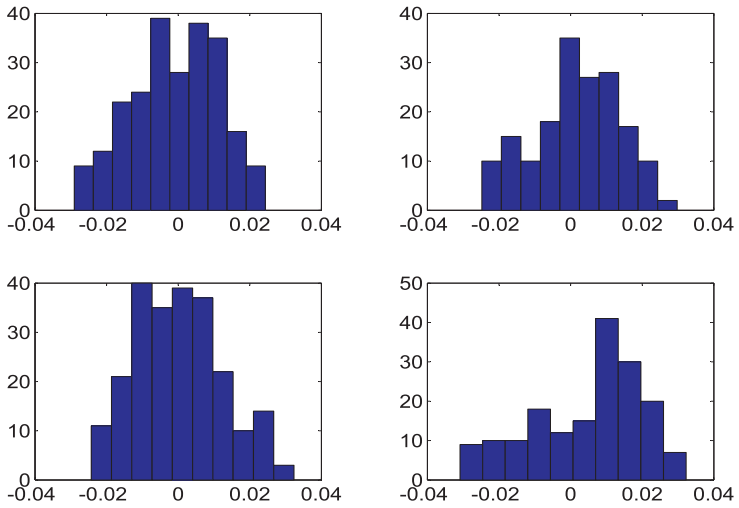

Figure 2.25: Forecasting errors in the random walk model.

The forecasts for 4-months ahead are described in table 2.12. Here the trend of the 1-month and 2-months forecasts is confirmed: The random walk gives good results for short times-to-maturity but is biased for long times-to-maturity (by almost 100 basis points). The semiparametric model has the highest bias for all times-to-maturity. The static Heston is also strongly biased for small times-to-maturity and the bias for longer times is also bigger than in the Heston and Nelson-Siegel approaches. These models have a quite small bias for l year-to-maturity. The biases in the Heston model seem to be smaller than in the Nelson-Siegel approach. The variation of the errors is smallest for the random walk and highest for the semiparametric approach. The other models show a similar variation in the errors.

2.4.4 Conclusion

We have analyzed the modeling and forecasting of variance swap curves. Reparametrizing the Heston model we consider a short Nelson-Siegel framework with two factors – level and slope. Generalizing this approach we analyzed also a full Nelson-Siegel model with the three factors – level, slope and curvature. Moreover, we considered a three factor semiparametric model. We analyzed the in-sample and out-of-sample performance of these models and compared the results to two benchmark models, the random walk and the static Heston model.

The in-sample fit of the Heston, the Nelson-Siegel and the semiparamet-

ric model are all satisfactory because the bias is small (less than 10 basis points). The variation of the errors measured by the interquartile range is also acceptably small (less than 20 basis points). The semiparametric model can be regarded as the best model with respect to the in-sample fit. Moreover, its nonparametric factors have the same interpretation – level, slope and curvature – as the factors in the parametric models.

We forecast the variance swap curves in these three models by forecasting the factor loadings. For the out-of-sample analysis we consider in addition the random walk and the static Heston model. The random walk gives the best results for short time ahead forecasts. The forecasts over longer periods show an increasing bias for long times-to-maturity in the random walk model. These forecasts have smaller errors in some models. The semiparametric approach has the worst out-of-sample performance. The static Heston model also seems to be inferior to the dynamic Heston or Nelson-Siegel model. These two models have forecasting power over long periods for long times-to-maturity and the Heston model seems slightly superior to the Nelson-Siegel parametrization.

We conclude that variance swap curves can be forecasted for long times-to-maturity. Such forecasts based on parametric models outperform the random walk. The random walk is good for short time ahead forecasts because of its perfect in-sample fit. Moreover, it gives good results for the short times-to-maturity because the variance swap curves are quite volatile and hard to model in this region. Semiparametric models give good in-sample fits but their out-of-sample performance appears inferior. The number of factors for parametric modelling cannot be determined in general. But the two-factor model slightly outperforms the three parameter generalization out-of-sample while the three factor approach gives of course the better in-sample fit.

Chapter 3

Option Pricing Models

Since the crash of 1987, implied volatilities are not constant as assumed in the Black-Scholes model but increase for smaller strikes. This phenomenon triggered the development of more complex models in order capture this and other important patterns of implied volatility surfaces.

Models for option pricing focused for a long time on fitting accurately observed prices of European options. Dupire showed that a perfect fit is possible in a nonparametric framework. But it is preferable in general to use parametric models where the parameters have economic interpretations. Moreover, researchers in academia and industry alike consider more and more the out-of-sample performance, i.e. good models should show reasonable dynamics of implied volatility surfaces. The direct control of these dynamics by parameters allows traders to incorporate directly their opinion on future movements. In addition, models are often calibrated to prices of variance swaps because these products are used for hedging and their markets are sufficiently liquid. As it is difficult for a single model to have all the described qualities there is a variety of models and each serves different products and different market regimes. In this chapter, we present some prominent option pricing models and illustrate some their advantages and disadvantages.

In section 3.1, we discuss some models popular in equity derivative markets and describe a semiparametric Heston model that replicates perfectly observed variance swap prices. In section 3.2, we present typical numerical techniques for the valuation of options considering the Heston model as an example.

3.1 Models

Black and Scholes showed how to price European options by replication. This idea of holding a portfolio of simple products whose value coincides with the value of the derivative is still fundamental to option pricing. Moreover, it leads to the idea of hedging and is thus applied by traders all the time.

Besides this seminal approach, they used a model of Samuelson where the returns have identical independent normal distributions and the stock S_t follows a geometric Brownian motion:

$$\frac{dS_t}{S_t} = rdt + \sigma dW_t$$

Although this model differs from reality in essential aspects it is still important because prices are quoted in volatilities implied by this model. Moreover, this model represents the intersection of more advanced models and it helps explaining phenomena in an intuitive way that are difficult to understand in complex models. In the following, we present some popular models that generalize this Black-Scholes model.

3.1.1 Local Volatility Models

As described in section 2.3, the implied volatility surface has a term structure. If the implied volatilities of at-the-money options on the DAX are 20% for 1 year to maturity and 22% for 2 years to maturity then clearly volatility is time dependent and this can be modelled in the Black-Scholes framework by making the volatility σ a function of time $\sigma(t)$, see Merton (1973).

But we saw in section 2.3 also that implied volatilities depend on the strike level. Hence, it is quite natural to extend Merton's time dependent volatility model to

$$\frac{dS_t}{S_t} = rdt + \sigma(t, S_t)dW_t \tag{3.1}$$

where the instantaneous volatility $\sigma(t, S_t)$ depends also on the spot level. Such models are specified by the deterministic function σ that determines the volatility only locally at (t, S_t). Hence, these models are called local volatility models.

Cox (1996) considered the parametric local volatility model

$$\frac{dS_t}{S_t} = rdt + \sigma S_t^{\beta-1} dW_t$$

where $\beta \in [0.5, 1]$. It is called the constant elasticity of variance model. For $\beta = 1$ this model is the original approach of Black and Scholes. Because of the few parameters this model has problems in replicating the implied volatility surfaces observed in reality. Moreover, the probability of reaching zero is positive if $\beta < 1$ and the process stays at zero once it hits this level. Hence, this model has some practical and theoretical shortcomings for equity indices that never touch zero.

Dupire (1994) left the parametric framework and showed that there exists a unique local volatility function σ that replicates perfectly a given implied volatility surface. An implied volatility surface represents the prices of European call options with payoff profile $f(S_T) = (S_T - K)^+$. Hence, the payoff profile satisfies

$$\frac{\partial f}{\partial K}(S_T) = -\mathbf{1}_{\{S_T > K\}} \text{ and } \frac{\partial^2 f}{\partial K^2}(S_T) = \delta_K(S_T).$$

Because of these special properties of the payoff profiles of European calls we can derive the (risk neutral) density $q(\cdot, T)$ of S_T from call prices $C(K, T)$. These prices and the density are related by

$$C(K,T) = \int \exp(-rT) q(x,T)(x-K)^+ dx.$$

Hence, the density can be calculated by taking derivatives

$$\begin{aligned}
\frac{\partial C}{\partial K}(K,T) &= -\exp(-rT) \int_K^\infty q(x,T)dx \\
\frac{\partial^2 C}{\partial K^2}(K,T) &= \exp(-rT) q(K,T)
\end{aligned}$$

This is the so-called Breeden-Litzenberger formula, see Breeden and Litzenberger (1978).

Moreover, the density q satisfies the Fokker-Planck equation, i.e.

$$\begin{aligned}
\frac{\partial \tilde{C}}{\partial T}(K,T) &= \int_K^\infty \frac{\partial q}{\partial T}(x,T)(x-K)dx \\
&= \int_K^\infty \left[\frac{1}{2} \frac{\partial^2 \{x^2 \sigma(T,x)^2 q(x,T)\}}{\partial x^2} - \frac{\partial \{xrq(x,T)\}}{\partial x} \right] (x-K)dx
\end{aligned}$$

where $\tilde{C}(K,T) = \exp(rT)C(K,T)$ denotes the undiscounted call prices. Partial integration yields

$$\begin{aligned}
\frac{\partial \tilde{C}}{\partial T}(K,T) &= \frac{1}{2}\sigma(T,K)^2 K^2 q(K,T) + \int_K^\infty rxq(x,T)dx \\
&= \frac{1}{2}\sigma(T,K)^2 K^2 \frac{\partial^2 \tilde{C}}{\partial K^2}(K,T) + r\left\{ \tilde{C}(K,T) - K\frac{\partial \tilde{C}}{\partial K}(K,T) \right\}.
\end{aligned}$$

Hence, the formula of Dupire holds

$$\sigma^2(T,K) = 2\frac{\frac{\partial \tilde{C}}{\partial T}(K,T) - r\{\tilde{C}(K,T) - K\frac{\partial \tilde{C}}{\partial K}(K,T)\}}{K^2\frac{\partial^2 \tilde{C}}{\partial K^2}(K,T)}. \tag{3.2}$$

Via the formula of Dupire a local volatility function is defined such that the corresponding diffusion model (3.1) for the stock price leads exactly to the option prices observed on the market. These prices are the input data for computing the local volatility function via (3.2). The perfect fit of the model is achieved by using a nonparametric local volatility function. As there are only a finite number of observed prices this nonparametric approach often requires interpolation or smoothing of the prices in order to get a surface. Another problem that is often encountered in nonparametric models lies in the dynamics of the prices. The forecasted implied volatility surfaces normally become flat in this model and do not show skew or smile patterns even if the surface that the model is calibrated to has a pronounced smile. This flattening of the implied volatility surfaces limits the applications of Dupire's model. Forward start options that depend only on the prices at the starting time cannot be priced realistically in Dupire's model. Hence, the same is true for cliquet options. But other less exotic derivatives like barrier options can be priced and hedged quite well in Dupire's model.

3.1.2 Stochastic Volatility Models

In section 2.4 we analyzed variance swaps, i.e. forward contracts on realized variance. These products are traded because the realized volatility of traded assets shows an economically significant variation over time. As this variation is not deterministic it seems natural to consider also in option pricing models a stochastic volatility in order to price and hedge options appropriately. Such stochastic volatility models are normally implemented via

$$\frac{dS_t}{S_t} = rdt + \sigma_t dW_t \tag{3.3}$$

where a stochastic process (σ_t) models the instantaneous volatility. The stochastic nature of the instantaneous volatility is underlined by figure 3.1 that shows the historical realized volatility for sample periods of different lengths.

We start with the stochastic volatility model of Heston (1993) because it is one of the most popular models in equity. It is given by the stochastic

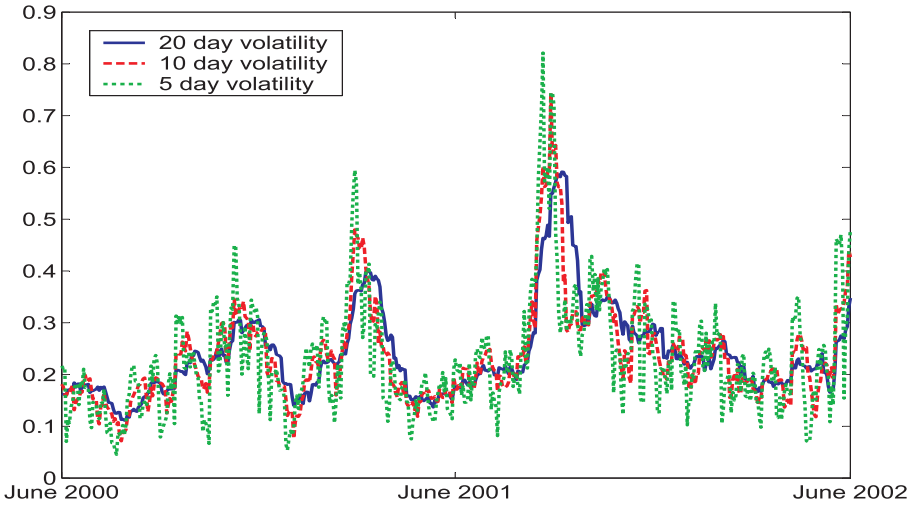

Figure 3.1: DAX realized volatility.

differential equations:

$$\frac{dS_t}{S_t} = rdt + \sqrt{V_t}dW_t^1 \tag{3.4}$$

where the variance process is modelled by a square-root process:

$$dV_t = \xi(\eta - V_t)dt + \theta\sqrt{V_t}dW_t^2 \tag{3.5}$$

and W^1 and W^2 are Wiener processes with correlation ρ. Here r denotes the risk free interest rate. The first equation models the stock returns by normal innovations with stochastic variance and the second equation models the instantaneous stochastic variance process as a square-root diffusion.

The parameters of the model all have economic interpretations: η is called the long variance because the process returns to this level. If the variance V_t is e.g. below the long variance then $\eta - V_t$ is positive and the drift drives the variance in the direction of the long variance. ξ controls the speed at which the variance is driven to the long variance. In calibrations, this parameter changes a lot and makes also the other parameters instable. To avoid this problem, the reversion speed is kept fixed in general. The volatility of variance θ controls mainly the kurtosis of the distribution of the variance. Moreover, there is an initial variance V_0 and a correlation ρ between the Brownian motions. The correlation is in general negative and models the leverage effect: When the stock price goes down then the variance goes up and viceversa. The parameters also control different aspects of the implied

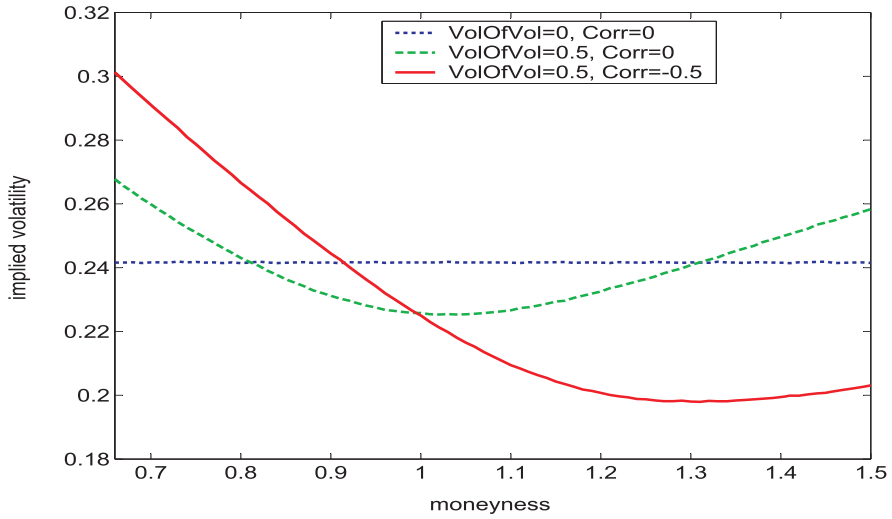

Figure 3.2: Effects of volatility of variance and correlation in the Heston model on the 1y implied volatility smile. (The other parameter are $\xi = 1, \eta = 0.3^2$ and $V_0 = 0.2^2$.)

volatility surface. The short (long) variance determines the level of implied volatility for short (long) maturities. The correlation creates the skew effect and the volatility of variance controls the smile. These effects are illustrated in figures 3.2 and 3.3.

As the variance is a stochastic process in this model, it is interesting and important to analyze how the process evolves. To this end, we present in figure 3.4 distributions of the volatility at future dates. In the Heston model there are two situations: If the volatility of variance is small enough (i.e. $\theta^2 < 2\xi\eta$) then the variance stays always strictly positive and the distributions of future volatility is concentrated around the long variance. But if the volatility of variance is too big (i.e. $\theta^2 > 2\xi\eta$) then the variance can touch zero and this is reflected in left skewed density of future volatility. This skew implies quite a high probability of a low volatility which does not seem realistic in all market regimes. Unfortunately, calibrations lead very often to parameters with a big volatility of variance.

Besides the statistical properties of a model numerical issues are of fundamental practical importance. In section 3.2, we discuss for the Heston model the pricing of plain vanilla options that is essential for model calibration. A very appealing feature of the Heston model is that European calls and puts can be priced in fast way by Fourier inversion techniques.

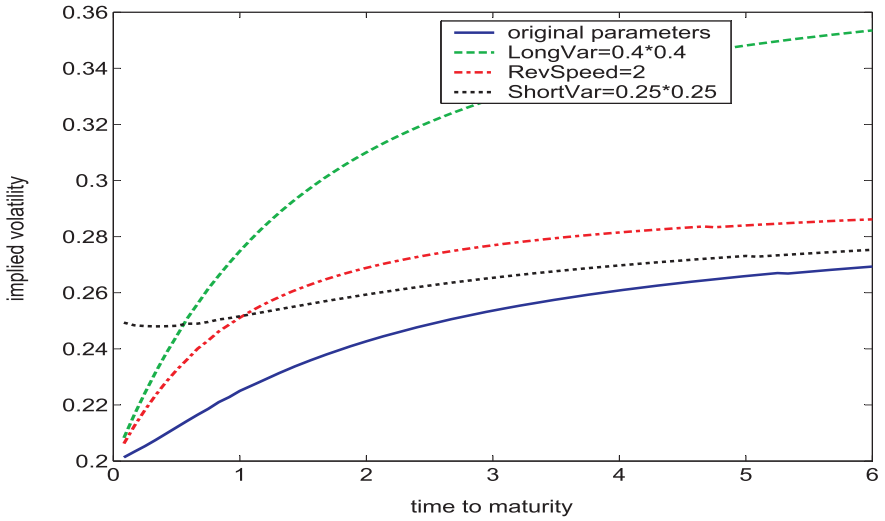

Figure 3.3: Effects of short variance, long variance and reversion speed in the Heston model on the ATM term structure of implied volatilities. (The parameters are $\xi = 1, \eta = 0.3^2, \theta = 0.5, \rho = -0.5$ and $V_0 = 0.2^2$.)

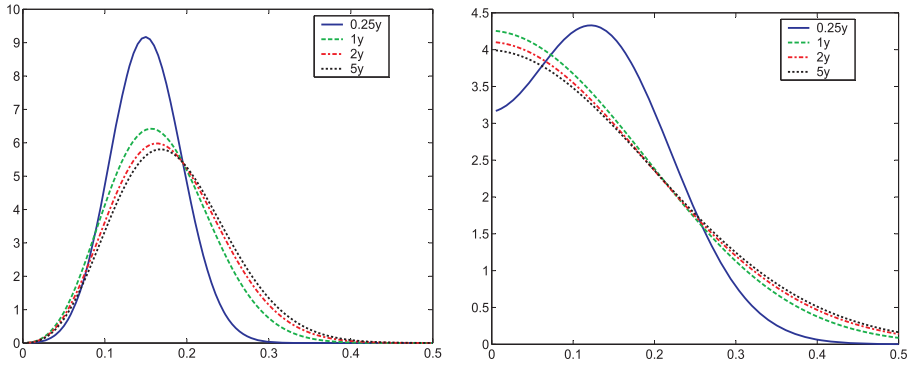

Figure 3.4: Probability density function of the volatility in the Heston model for the parameters $\xi = 1, \eta = 0.2^2, \rho = -0.5$ and $V_0 = 0.15^2$. (left: $\theta = 0.2$, right: $\theta = 0.4$)

The model of Hull and White (1987) that can be regarded as the first solved stochastic volatility model has no such semiclosed solution. Instead its solution can be represented by a Taylor series. The approach models the instantaneous variance process by a geometric Brownian motion

$$\begin{aligned}\frac{dS_t}{S_t} &= r dt + \sqrt{V_t} dW_t^1 \\ \frac{dV_t}{V_t} &= \zeta dt + \theta dW_t^2\end{aligned}$$

where W^1 and W^2 are uncorrelated Wiener processes. As the instantaneous variance has a lognormal distribution the volatility $\sigma_t = \sqrt{V_t}$ is always positive and can be represented by

$$\sigma_t = \sigma_0 \exp\left\{\frac{1}{2}\theta W_t^2 + \frac{1}{2}(\zeta - \frac{1}{2}\theta^2)t\right\}$$

We show corresponding densities for different times t in figure 3.5 that illustrates how the volatility tends to zero over time. Calibrations to market data lead in most cases to $\zeta < 0$. This implies that the variance is zero reverting. Moreover, the expectation of volatility and the mode of its density converge to zero over time. These effects are illustrated in figure 3.5. But these theoretical properties of the volatility are observed on the markets only rarely. This shortcoming of the model and the not modelled leverage effect make the model less popular in practice. Moreover, if the model is generalized by allowing the Brownian motions that drive the stock and variance processes to be correlated then the prices of European options cannot be represented by a Taylor series.

The stochastic volatility model of Schöbel and Zhu (1999) allows for a correlation between the two processes. But in contrast to the Hull/White model European options can be priced efficiently by Fourier inversion techniques. The model of Schöbel and Zhu (1999) is given by

$$\begin{aligned}\frac{dS_t}{S_t} &= r dt + \sigma_t dW_t^1 \\ d\sigma_t &= \xi(\eta - \sigma_t)dt + \theta dW_t^2\end{aligned}$$

where W^1 and W^2 are Wiener processes with correlation ρ. This model generalizes the model of Stein and Stein (1991) that has no correlation between the Brownian motions. The volatility that is modelled directly is given by an Ornstein-Uhlenbeck process. Hence, the volatility has a normal distribution and negative values have a positive probability. The sign of the volatility

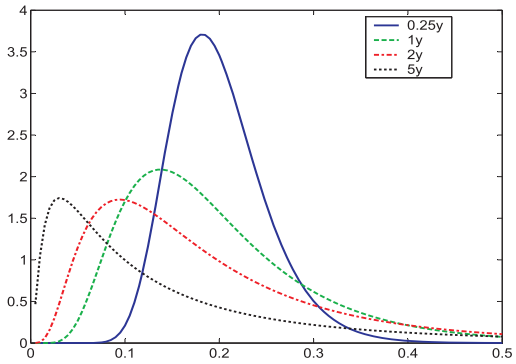

Figure 3.5: Probability density function of the volatility in the Hull-White model for the parameters $\zeta = 0, \theta = 1$ and $V_0 = 0.2^2$.

bears meaning as a sign modifier of the correlation. As the leverage effect is rather constant over time the correlation should be negative. Thus, the changing sign of the correlation is rather unrealistic. Moreover, a negative volatility is not intuitive. For these reasons, the model is not popular in practice. In addition, the distribution of volatility in this model is similar to the distribution of volatility in the Heston model when the volatility of variance is big.

In the three described stochastic volatility models, the volatility often tends take quite low values in the future. The model of Scott (1987) implies a stable and rather realistic evolution of volatility. The model is given by

$$\begin{aligned} \frac{dS_t}{S_t} &= rdt + \exp(y_t)dW_t^1 \\ dy_t &= \xi(\eta - y_t)dt + \theta dW_t^2 \end{aligned}$$

where W^1 and W^2 are Wiener processes with correlation ρ. The volatility $\sigma_t = \exp(y_t)$ is the exponential of an Ornstein-Uhlenbeck process. As the distribution of the Ornstein-Uhlenbeck process y converges to normal distribution the volatility has a stable lognormal distribution. This is illustrated in figure 3.6 that shows the densities of the volatility for different future times. But in contrast to this intuitive dynamics of the volatility prices of European options have to be computed by extensive Monte-Carlo simulations and the fit of the model to plain vanilla prices is not satisfactory.

The four stochastic volatility approaches described so far model the stock

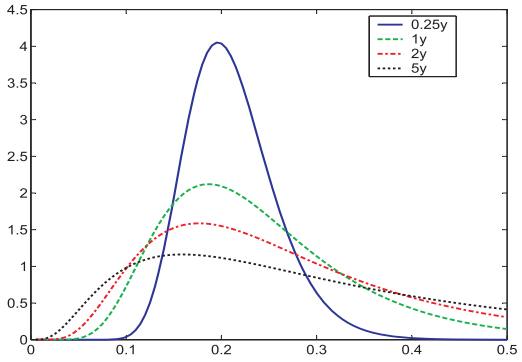

Figure 3.6: Probability density function of the volatility in the Scott model for the parameters $\sigma_0 = 0.3, \theta = \log 0.2, \chi = 0.1, \alpha = 0.4, \alpha^2/2\chi = 2$.

Model	Correlation	$f(V_t)$	(V_t)
Hull-White	$\rho = 0$	$\sqrt{V_t}$	geometric Brownian motion
Scott	$\rho = 0$	$\exp(V_t)$	Ornstein-Uhlenbeck process
Heston	$\rho \neq 0$	$\sqrt{V_t}$	square-root process
Schöbel-Zhou	$\rho \neq 0$	V_t	Ornstein-Uhlenbeck process

Table 3.1: Summary of stochastic volatility models.

price process by

$$\frac{dS_t}{S_t} = rdt + f(V_t)dW_t \tag{3.6}$$

with different functions f. Moreover, the models use different processes (V_t) for the variance/volatility. In this generic framework the four models are summarized in table 3.1.

A stochastic volatility model that does not fit directly into this framework (3.6) was introduced by Hagan et al. (2002) and is called the SABR model

$$\begin{aligned} \frac{dS_t}{S_t} &= rdt + S_t^{\beta-1}\alpha_t dW_t^1 \\ d\alpha_t &= \nu\alpha_t dW_t^2 \end{aligned}$$

where W^1 and W^2 are Wiener processes with correlation ρ. The parameters $\beta \in [0.5, 1]$ and ν control the behaviour of the model: The constant elasticity

of variance model of section 3.1.1 corresponds to $\nu = 0$ and for $\beta = 1$ the model is quite similar to the Hull-White model. Hence, this model can be regarded as a mixture of different models. An appealing feature of the model is that implied volatilities can be approximated in a simple way. Moreover, the model implies quite realistic dynamics of the implied volatility surfaces. But for $\beta < 1$ the stock price can be absorbed at zero as in the constant elasticity of variance model. This is rather unrealistic for equity indices.

3.1.3 Lévy Models

The local volatility models of section 3.1.1 and the stochastic volatility models of section 3.1.2 are pure diffusion models where the evolution is driven by normal innovations. Adding jump components in form of a compound Poisson process makes these models jump-diffusions where the innovations are not normal anymore. We consider examples of these finite activity Lévy models where there are only a finite number of jumps in every finite time interval. Moreover, we discuss pure Lévy models of infinite activity that are derived by time changed Brownian motions.

Merton (1976) extended the Black-Scholes by adding normally distributed jumps to the returns:

$$\frac{dS_t}{S_{t-}} = rdt + \sigma dW_t + dJ_t$$

where W is a Brownian motion and the process J is independent of W with

$$J_t = \sum_{j=1}^{N_t} (Y_j - 1)$$

for a Poisson process N and independent log normally distributed random variables $Y_1, Y_2, \ldots$. The Y_j are the ratios of the asset price before and after the j-th jump. Hence, they are modelled by positive variables. The resulting distribution of the asset price can be represented by a Poisson mixture of lognormal distributions because the Poisson process is independent of the Brownian motion. This allows also to represent the price of European call options as a series where Black-Scholes prices are weighted by Poisson probabilities. The model with its few parameters does not fit market data well but it creates a smile. An appealing feature of the model is that this smile does not disappear over time but is constant on the moneyness scale. But the implied volatility surfaces become flat for long times to maturity. These smile properties are shared by all exponential Lévy models. The Merton model and

the Heston model have been combined by Bates (1996). This mixed model has been further extended by Duffie et al. (2000) by considering jumps also in the variance process.

Besides such finite activity Lévy models, there are also approaches with an infinite number of small jumps. Madan and Seneta (1990) introduced the Variance Gamma process for stock returns and later analyzed option pricing in this model. This model can be derived by subordination, i.e. time change of a Lévy process. Such a time change is interpreted as business time. It reflects economically the random arrival of the information that drives the evolution of the asset price. The time change is modelled in this approach by a Gamma process, i.e. a Lévy process (T_t) with $\mathcal{L}(T_1) = Gam(1, \nu)$. As Gamma random variables can take on only positive values time increases always and thus T is a subordinator, i.e. an increasing Lévy process. This time change is applied to a Brownian motion with drift θ and volatility σ

$$b_t = \theta t + \sigma W_t.$$

The resulting time changed process $X_t := b_{T_t}$ is called a Variance Gamma process because the variance of X_t depends on the Gamma process T_t. In the Variance Gamma model for asset prices, the logarithm of the asset price (S_t) is modelled by a Variance Gamma process. The discounted price process becomes a martingale in this model under a special drift

$$S_t = S_0 \exp\left\{ rt + \frac{t}{\nu} \log\left(1 - \theta\nu - \frac{\sigma^2\nu}{2}\right) + X_t \right\}.$$

Hence, options can be priced under this risk neutral measure. The prices of European options can be represented as double integrals corresponding to the time change and the Brownian motion. Thus these prices are given by an integral over Black-Scholes prices.

The Variance Gamma has been generalized by Carr et al. (2002). In the generalized approach that is known as CGMY model, the time changed Brownian motion is replaced by a Lévy process with Lévy density

$$k(x) = \begin{cases} C\frac{\exp(-G|x|)}{|x|^{1+Y}}, & x < 0 \\ C\frac{\exp(-M|x|)}{|x|^{1+Y}}, & x > 0 \end{cases}$$

where $C > 0, G \geq 0, M \geq 0$ and $Y < 2$. The parameter C controls the overall level of activity while the parameter Y characterizes the fine structure of the process. The parameters G and M control the rate of decay on the right and left of the Lévy density. The Variance Gamma process is recovered by $Y = 0$.

Both infinite activity asset price models have appealing economic interpretations of the parameters and can replicate the smile well. However, the term structure of implied volatility surfaces is hard to fit. Moreover, the approaches seem technically more demanding because of the concepts of time change of Lévy processes. For these reasons, the models do not seem to be very popular in industry at this time.

3.1.4 Market Models

Besides the above approaches that model only the underlying price process and perhaps an unobservable stochastic volatility there are approaches that model also the prices of liquid options like European options or variance swaps. In these market models the prices of liquid options are not derived from the underlying but are regarded as idiosyncratic processes. The difficulty of these approaches lies in guaranteeing that no arbitrage opportunities arise.

Schönbucher (1999) introduced such a market model for equities by modelling the prices of European options directly by the corresponding implied volatilities. In this approach, the underlying is modelled by

$$dS_t = rS_t dt + \sigma S_t dW_t^0$$

where σ denotes the stochastic actual volatility. Moreover, the price of a European option with strike K and maturity T is modelled via its implied volatility $\hat{\sigma}(T,K)$ that is assumed to follow

$$d\hat{\sigma}(T,K) = u(T,K)dt + \gamma(T,K)dW^0 + v(T,K)dW$$

where W is an N-dimensional Brownian motion and v a corresponding volatility vector.

We illustrated this market model by considering one implied volatility only. The dynamics of the corresponding option price C are given by Itô's formula

$$dC = C_t dt + C_S dS + \frac{1}{2}\sigma^2 S^2 C_{SS} dt + C_{\hat{\sigma}} d\hat{\sigma} + \frac{1}{2} C_{\hat{\sigma}\hat{\sigma}} d\hat{\sigma} d\hat{\sigma} + C_{\hat{\sigma}S} d\hat{\sigma} dS.$$

This representation applied to the no arbitrage condition $rCdt = \mathrm{E}(dC)$ gives the no arbitrage drift

$$u = \frac{1}{2C_{\hat{\sigma}}} \left\{ (\hat{\sigma}^2 - \sigma^2) S^2 C_{SS} - C_{\hat{\sigma}\hat{\sigma}} v^2 - 2\gamma\sigma S C_{\hat{\sigma}S} \right\}$$

using the Black-Scholes partial differential equation. As the greeks are known for European options in the Black Scholes model, the no arbitrage drift restriction can be simplified to

$$\hat{\sigma}u = \frac{1}{2\tau}(\hat{\sigma}^2 - \sigma^2) - \frac{1}{2}d_+d_-v^2 + \frac{d_-}{\sqrt{\tau}}\sigma\gamma$$

where $d_\pm = \{\log(S_t/K) + (r \pm \frac{1}{2}\hat{\sigma}^2)(T-T)\}/(\hat{\sigma}\sqrt{T-t})$. So-called volatility bubbles that are comparable to arbitrage and can emerge for $t \to T$ can be avoided by a further constraint on the spot volatility, see Schönbucher (1999) for details.

The described method for one option can be generalized to a continuum of option prices. Bühler (2006) followed a similar idea by modelling the underlying and the variance swap curve directly by stochastic processes. Both approaches turn out to be rather complicated when a finite number of prices are modelled directly. On the other hand, the continuous approach faces the problem that only a finite number of options or variance swaps are traded.

3.1.5 A Semiparametric Stochastic Volatility Model

In this section, we propose a model for equity derivatives pricing that combines a stochastic variance structure with a nonparametric long variance function. This semiparametric specification allows a perfect replication of variance swap prices and a fast calibration to plain vanilla prices by characteristic function techniques. Hence, it provides a good fit to market prices, is numerically trackable and yields moreover reasonable forward implied volatility surfaces.

Motivation

In recent years, the popularity of forward structures and the growth of variance swap markets motivated the search for new option pricing models. These two impulses drew the attention of researchers in academia and industry to the evolution of the price surfaces and the fit to variance swap curves.

Hence, a good model should fulfill nowadays different requirements: Besides a good in-sample fit to prices of European options, the model is expected to replicate the term structure presented by the prices of variance swaps. Moreover, the out-of-sample performance is essential for products like cliquet options. In addition to these requirements, parsimonous parametrizations are preferred where each parameter can be interpreted in an economic

way. Finally, the model must be tractable from the numerical point of view in order to allow efficient calibrations to market data.

Dupire (1994) considered a one factor model

$$dS_t/S_t = rdt + \sigma(t, S_t)dB_t$$

and showed that there is a unique local volatility function σ implied by the market that fits perfectly all the prices of European call and put options. As the model has only one factor it is numerically relatively easy. The perfect fit is due to the purely nonparametric structure. Like in many other purely nonparametric methods the perfect fit comes at the cost of an unrealistic evolution of the price surfaces. Another approach are two factor models where the volatility is modeled explicitly by a stochastic process. Heston (1993) considered in this class a square-root process for the variance

$$\begin{aligned} dS_t/S_t &= rdt + \sqrt{\zeta_t}dB_t \\ d\zeta_t &= \kappa(\eta - \zeta_t)dt + \theta\sqrt{\zeta_t}dW_t \end{aligned}$$

This model is numerically more difficult because of the two factors. Its fit to market data is not always satisfactory because the term structure can not be replicated well. The model is purely parametric and all its five parameters have an economic meaning. Moreover, its out-of-sample performance is clearly superior to the local volatility model. In order to combine the good features of these two models mixed models have been considered, e.g.

$$\begin{aligned} dS_t/S_t &= rdt + \sigma(t, S_t)\sqrt{\zeta_t}dB_t \\ d\zeta_t &= \kappa(\eta - \zeta_t)dt + \theta\sqrt{\zeta_t}dW_t \end{aligned}$$

But this approach is overparametrized so that a stable calibration is rather difficult.

Instead of adding the model characteristics we propose a semiparametric model that has the parametric structure of the Heston model. We make the parameter η a deterministic function depending on time

$$\begin{aligned} dS_t/S_t &= rdt + \sqrt{\zeta_t}dB_t \\ d\zeta_t &= \kappa(\eta_t - \zeta_t)dt + \theta\sqrt{\zeta_t}dW_t \end{aligned}$$

In this way, our model is "less" nonparametric than the local volatility model whose nonparametric function is two dimensional. But at the same time we can fit perfectly the term structure of implied volatility surfaces represented

by variance swap prices, see 3.8. Moreover, the model is numerically still tractable because the prices of European options can be computed by fast Fourier transform methods. As the model implies also reasonable evolutions of the volatility surfaces it combines good features of the local volatility and the Heston model while keeping the model dimension at same time relatively low.

Semiparametric modelling

First, we review the Heston model and show that plain vanilla prices can still be computed by characteristic function techniques if the long variance is a function of time. Then we compute the prices of variance swaps in such a model and identify conditions for the long variance function that imply a perfect fit to variance swap prices.

The original model of Heston was described in section 3.1.2. It has a constant long variance η. But this constraint can be relaxed by considering parametric functions of time for the long variance. A popular parametrization is

$$\eta_t = m + (\eta_0 - m)e^{-ct}.$$

This form replaces the long variance parameter by two additional parameter so that this version of the Heston model has five parameters if the mean reversion speed is fixed. These additional parameters make the calibration more difficult and time consuming. Although the fit to market data is better it is in general far from the perfect fit of the local volatility model. Hence, this model improves the fit without attaining a perfect fit to any prices.

We propose to make the long variance a nonparametric function η_t. We calibrate it (and the initial variance) to variance swap price in such a way that these prices are replicated perfectly. In addition, we are left afterwards with a simple calibration of a two parameter model to plain vanilla prices. This calibration to plain vanilla prices can be implemented efficiently because these prices can be still computed by the Fourier transform approach. This method is based on the characteristic function of the logarithm of the stock price and this function can be still computed for a time dependent long variance function.

We show next how to compute this characteristic function

$$\psi_T(z) = \mathrm{E}\{\exp(\mathbf{i}z \log S_T)\}.$$

The process $dB_t^z = -\mathbf{i}z\sqrt{\zeta_t}dt + dB_t$ is a Brownian motion under the measure P^z with density $dP^z/dP = \exp(\mathbf{i}z\int_0^T \sqrt{\zeta_t}dB_t + z^2/2\int_0^T \zeta_t dt)$. Under P^z, the characteristic function can be represented as

$$\begin{aligned}\psi_T(z) &= e^{\mathbf{i}z(rT+\log S_0)}\,\mathrm{E}\left\{\exp\left(\mathbf{i}z\int_0^T \sqrt{\zeta_t}dB_t - \mathbf{i}z/2\int_0^T \zeta_t dt\right)\right\}\\ &= e^{\mathbf{i}z(rT+\log S_0)}\,\mathrm{E}^z\left\{\exp\left[-(\mathbf{i}z+z^2)/2\int_0^T \zeta_t dt\right]\right\}\end{aligned}$$

Moreover, the process (ζ_t) can be represented as

$$d\zeta_t = \tilde{\kappa}(\tilde{\eta}_t - \zeta_t)dt + \theta\sqrt{\zeta_t}dW_t^z$$

where $\tilde{\kappa} := \kappa - \rho\mathbf{i}z\theta$, $\tilde{\eta}_t := \eta_t\kappa/\tilde{\kappa}$ and W^z is a Brownian motion under P^z with correlation ρ with respect to B^z.

Hence, the characteristic function ψ_T is determined by the Laplace transform of the integral of a square-root process $dx_t = (m_t - kx_t)dt + \xi\sqrt{x_t}dW_t$. The joint transform of $(x_T, \int x_t dt)$ can be derived by considering a martingale associated to (x_t) and using the Itô formula. This approach leads to a partial differential equation that can be solved explicitly even when the long term equilibrium level is time dependent. This solution has the form

$$\mathrm{E}(e^{-\mu x_T - \lambda\int_0^T x_t dt}) = e^{-x_0 A_T(\mu,\lambda) - B_T(\mu,\lambda)}$$

where the functions A_T and B_T satisfy

$$A_T(\mu,\lambda) = \frac{\alpha + ae^{\gamma t}}{\beta + be^{\gamma t}}$$

and

$$B_T(\mu,\lambda) = \int_0^T m_{T-t}A_t(\mu,\lambda)dt$$

with the constants $\alpha := \lambda(\gamma+b)-2\mu$, $a := \lambda(\gamma-b)+2\mu$, $\beta = -\xi^2\lambda+\gamma+b$, $b := \xi^2\lambda + \gamma - b$ and $\gamma := -\sqrt{k^2+\xi^2\mu}$.

From this characteristic function ψ the prices of plain vanilla options can be computed by a method of Carr and Madan (1998). The price of a European call option with maturity T and strike K is then given by

$$\frac{\exp(-\alpha k)}{\pi}\int_0^\infty e^{-\mathbf{i}vk}\phi_T(v)dv$$

where $k := \log K$ is the log strike, $\alpha > 0$ a dampening factor and

$$\phi_T(v) = \frac{\psi_T\{k - \mathbf{i}(\alpha+1)\}}{(\mathbf{i}k+\alpha)(\mathbf{i}k+\alpha+1)}.$$

Hence, a nonparametric long variance function still allows a fast computation of plain vanilla prices. We choose this nonparametric function in such a way that the prices of variance swaps are perfectly replicated. This implies that the model fits perfectly the term structure.

A variance swap trades a realized variance against a fixed variance. We define the realized variance in the period $[0;T]$ with trading days $0 = t_0 < t_1 < \ldots < t_n = T$ and 252 trading days a year by

$$\frac{252}{n}\sum_{i=1}^{n}\left(\log\frac{S_{t_i}}{S_{t_{i-1}}}\right)^2.$$

This sum can be approximated by the quadratic variation of $\log S$

$$< \log S >_T \approx \sum_{i=1}^{n}\left(\log\frac{S_{t_i}}{S_{t_{i-1}}}\right)^2, \tag{3.7}$$

see e.g. Protter (2004).

Thus, the fair price of a (zero strike) variance swap is given by

$$\mathrm{E}\left(\frac{1}{T} < \log S >_T\right)$$

under a martingale measure that is used for pricing. In the following, we work with the price of realized variance that is not annualized

$$V(T) := \mathrm{E}(< \log S >_T).$$

Next, we compute the price of a variance swap in the Heston model with time dependent long variance and see what equations the initial variance ζ_0 and the long variance function η_t have to fulfill in order to fit perfectly an observed variance swap price curve.

Using the approximation (3.7) we identify the realized variance as $\int \zeta_t dt$ and its price as $\mathrm{E}(\int \zeta_t dt)$. In order to compute this price, we first analyze

$$\mathrm{E}(\zeta_t) = \zeta_0 + \int_0^t \kappa\{\eta_s - \mathrm{E}(\zeta_s)\}ds.$$

This expectation considered as a function of t fulfills an inhomogeneous linear ordinary differential equation and hence has the representation

$$\mathrm{E}(\zeta_t) = \exp(-\kappa t)\{\zeta_0 + \int_0^t \kappa\eta_s \exp(\kappa s)ds\}.$$

Thus, the price of (not annualized) realized variance is given by

$$\begin{aligned} V(T) &= \mathrm{E}\left(\int_0^T \zeta_t dt\right) \\ &= \int_0^T \exp(-\kappa t)\{\zeta_0 + \int_0^t \kappa\eta_s \exp(\kappa s)ds\}dt. \end{aligned}$$

By taking the derivative we see

$$V'(T) = \exp(-\kappa T)\{\zeta_0 + \int_0^T \kappa\eta_s \exp(\kappa s)\}.$$

Differentiating once again we obtain

$$\begin{aligned} \kappa\eta_T \exp(\kappa T) &= \frac{d}{dT}\{\exp(\kappa T)V'(T) - \zeta_0\} \\ &= \kappa \exp(\kappa T)V'(T) + \exp(\kappa T)V''(T) \end{aligned}$$

Hence, the model fits the variance swap curve perfectly if and only if

$$\begin{aligned} \eta_t &= V'(t) + \frac{1}{\kappa}V''(t),\ t > 0 \\ \zeta_0 &= V'(0). \end{aligned} \tag{3.8}$$

As in reality no variance swap curves are observed but only a finite number of prices, the curves have to be constructed by interpolation or smoothing. We prefer spline smoothing that we also applied in section 2.4. Alternatively, traders can specify these curves.

3.2 Option Valuation Techniques

In this section, we illustrate some standard numerical techniques for pricing options. To this end, we focus on the stochastic volatility model of Heston (1993) because of its popularity. But similar approaches can be applied in many of the models described in section 3.1.

3.2.1 Fourier Transforms

One reason for the popularity of the Heston model are the semiclosed price formulas for European options that describe the price by two Fourier integrals. Carr and Madan (1999) have generalized this approach further so that the prices are given by one Fourier integral. Moreover, their approach is applicable to a wide range of option pricing models where the characteristic function of the log price is given analytically.

The basic idea of the method is to develop an analytic expression for the Fourier transform of the option price and to get the price by Fourier inversion. As the Fourier transform and its inversion work for square-integrable functions according to Plancherel's theorem we do not consider directly the option price but a modification of it.

Let $C_T(k)$ denote the price of a European call option with maturity T and strike $K = \exp(k)$. Its value is then given by:

$$C_T(k) = \int_k^\infty e^{-rT}(e^s - e^k)q_T(s)ds$$

where q_T is a risk-neutral density of $s_T = \log S_T$. The function C_T is not square-integrable because $C_T(k)$ converges to S_0 for $k \to -\infty$. Hence, we consider the modified function:

$$c_T(k) = \exp(\alpha k)C_T(k)$$

which is square-integrable for a suitable $\alpha > 0$. The choice of α may depend on the model for (S_t). The Fourier transform of c_T is defined by:

$$\psi_T(v) = \int_{-\infty}^\infty e^{\mathbf{i}vk}c_T(k)dk.$$

The expression for ψ_T can be computed directly after an interchange of integrals:

$$\begin{aligned}
\psi_T(v) &= \int_{-\infty}^\infty e^{\mathbf{i}vk}\int_k^\infty e^{\alpha k}e^{-rT}(e^s - e^k)q_T(s)dsdk \\
&= \int_{-\infty}^\infty e^{-rT}q_T(s)\int_{-\infty}^s \{e^{\alpha k+s} - e^{(\alpha+1)k}\}e^{\mathbf{i}vk}dkds \\
&= \int_{-\infty}^\infty e^{-rT}q_T(s)\{\frac{e^{(\alpha+1+\mathbf{i}v)s}}{\alpha+\mathbf{i}v} - \frac{e^{(\alpha+1+\mathbf{i}v)s}}{\alpha+1+\mathbf{i}v}\}ds \\
&= \frac{e^{-rT}\phi_T\{v-(\alpha+1)\mathbf{i}\}}{\alpha^2+\alpha-v^2+\mathbf{i}(2\alpha+1)v}
\end{aligned}$$

where ϕ_T is the Fourier transform of q_T. A sufficient condition for c_T to be square-integrable is given by $\psi_T(0)$ being finite. This is equivalent to

$$E(S_T^{\alpha+1}) < \infty.$$

Now, we get the desired option price in terms of ψ_T by the Fourier inversion

$$C_T(k) = \frac{\exp(-\alpha k)}{\pi} \int_0^\infty e^{-\mathbf{i}vk} \psi(v) dv.$$

This integral can be computed numerically by:

$$C_T(k) \approx \frac{\exp(-\alpha k)}{\pi} \sum_{j=0}^{N-1} e^{-\mathbf{i}v_j k} \psi(v_j) \eta \tag{3.9}$$

where $v_j = \eta j,\ j = 0, \ldots, N-1$ and $\eta > 0$ is the distance of the points of the integration grid.
Formula (3.9) suggests to calculate the prices with the FFT which is an efficient algorithm for computing the sums

$$w_u = \sum_{j=0}^{N-1} e^{-\mathbf{i}\frac{2\pi}{N}ju} x_j, \text{ for } u = 0, \ldots, N-1 \tag{3.10}$$

In general, strikes near the spot price are of interest because such options are traded most frequently. We consider thus an equidistant spacing of the log strikes around the log spot price s_0:

$$k_u = -\frac{1}{2}N\zeta + \zeta u + s_0, \text{ for } u = 0, \ldots, N-1$$

where $\zeta > 0$ denotes the distance between the log strikes. Substituting these log strikes in the approximation yields for $u = 0, \ldots, N-1$

$$C_T(k_u) \approx \frac{\exp(-\alpha k)}{\pi} \sum_{j=0}^{N-1} e^{-\mathbf{i}\zeta\eta ju} e^{\mathbf{i}(\frac{1}{2}N\zeta - s_0)v_j} \psi(v_j)\eta.$$

Now, the FFT can be applied to

$$x_j = e^{\mathbf{i}(\frac{1}{2}N\zeta - s_0)v_j} \psi(v_j), \text{ for } j = 0, \ldots, N-1$$

provided that

$$\zeta\eta = \frac{2\pi}{N}.$$

This constraint however leads to the following trade-off: The parameter N controls the computation time and thus is often determined by the problem. So the right hand side may be regarded as given or fixed. One would like to choose a small ζ in order to get many prices for strikes near the spot price.

But the constraint implies then a big η giving a coarse grid for integration. So we face a trade-off between accuracy and the number of interesting strikes.

Chourdakis (2005) developed an option pricing approach based on the fractional Fourier transform where this constraint does not exist anymore. The method requires the evaluation of one fractional N-point FFT that is computationally equivalent to three $2N$-point FFTs. But this speed disadvantage is outbalanced by the fact that only a smaller number N is necessary for the same accuracy as in the traditional FFT approach.

If the Fourier transform methods are applied to calculate the price of European options for whole range of strikes then the prices deep out of the money become inaccurate. Because of this, the method is often only applied for calculating a single price. Another problem of the approach is given by the evaluation of complex logarithms because they introduce sometimes numerical problems leading to instability. These numerical problems have been analyzed in a number of studies, e.g. Kahl and Jäckel (2005) or Albrecher et al. (2007).

3.2.2 Monte Carlo Simulations

In the last section, we discussed a fast pricing method that works only for some models and special options. We consider now with Monte Carlo simulations an approach that is applicable to a much wider range of models and options. Despite its apparent simplicity, we illustrate some difficulties by the example of discretization errors in the Heston model (3.4).

Option prices are given by the expectation of the discounted payoff under a risk neutral measure. These expectations can be computed by Monte Carlo simulations because of the strong law of large numbers. To this end, we have to sample from the distribution of the stock price process. There exist several approaches for this simulation, e.g. changing coordinates, exact simulation or Euler discretizations.

If we want to change coordinates, we can consider instead of the variance process the corresponding volatility process. Its dynamics are given by Itô's formula

$$d\sqrt{V_t} = \left(\frac{\zeta\eta - \theta^2/4}{2\sqrt{V_t}} - \frac{1}{2}\zeta\sqrt{V_t} \right) dt + \frac{1}{2}\theta dW_t^2$$

Such transformations are helpful for different reasons, e.g. the resulting equation has a constant diffusion coefficient that is potentially useful for reducing

the discretization error. But the application of Itô's formula is valid only if the square root transformation is twice differentiable on the domain of the process. In practical problems of equity derivatives, the model parameters often make the origin attainable for the variance process. As the square root is not differentiable at the origin the transformation method turns out to be problematic for the Heston model.

Broadie and Kaya (2006) developed a method to simulate the Heston model without bias. But this exact sampling comes at the cost of high computation times. The approach requires sampling from a noncentral chi-squared distribution and from another distribution whose characteristic function is known. The second sampling is done by the transform method that requires the numerical inversion of the cumulative distribution function that can be computed by Fourier inversion. Hence, this second sampling is quite time consuming. After these samples have been drawn the logarithmic stock price has a normal distribution so that this sampling is simple. But because of the second sampling this method is quite time consuming and hence has only limited practical applications.

Probably the approach most often encountered in industry is the Euler discretization

$$V_{t+\Delta t} = f_1(V_t) + \zeta\{\eta - f_2(V_t)\}\Delta t + \theta\sqrt{f_3(V_t)}\Delta W_t^2$$
$$\log(S_{t+\Delta t}) = \log(S_t) + \{r - 0.5 f_4(V_t)\}\Delta t + \sqrt{f_5(V_t)}\Delta W_t^1$$

where the functions $f_1, \ldots, f_5$ are required to fulfill $f_i(x) = x,\ x \geq 0$ and $f_3, f_4, f_5 \geq 0$. (Alternatively, the not transformed stock price process can be discretized directly.) The well known absorption scheme is specified by $f_i(x) = x^+,\ i = 1, \ldots, 5$. As discussed in Lord et al. (2006) the absorption scheme turns out to be biased if the origin can be attained by the variance process. In order to avoid this shortcoming other schemes have been proposed, e.g. partial truncation by Deelstra and Delbaen (1998): $f_1(x) = f_2(x) = x, f_3(x) = \ldots = f_5(x) = x^+$. Lord et al. (2006) proof the strong convergence of their full truncation scheme ($f_1(x) = x, f_2(x) = \ldots = f_5(x) = x^+$). Moreover, they show that their approach leads to the smallest bias of the price of a European call.

3.2.3 Partial Differential Equations

In this section, we illustrate how to price options in the Heston model by partial differential equations. In contrast to the Black-Scholes model, the Heston model is not complete, i.e. we can replicate all option payoffs with the underlying and the bonds only. As European options are traded quite

liquidly we can use also these derivatives for hedging and then we are able replicate all payoff profiles perfectly.

Hence, we consider besides the stock price process (3.4) (with variance process (3.5)) a bond $B_t = \exp(rt)$ with constant interest rate r and an option whose price $c(t, S_t, V_t)$ depends on time, stock price and variance. We hedge an option with price $u(t, S_t, V_t)$ by a trading strategy that consists of α_t bonds, β_t stocks and γ_t option at time t. Thus, the value of the hedge portfolio is

$$h_t := \alpha_t B_t + \beta_t S_t + \gamma_t c(t, S_t, V_t)$$

We require the hedging strategy to be self financing, i.e. $dh_t = \alpha_t dM_t + \beta_t dS_t + \gamma_t dc(t, S_t, V_t)$ and to replicate the option, i.e. $u(t, S_t, V_t) = h_t$. Hence, the increments $du(t, S_t, V_t)$ and dh_t are also equal. These expressions can be computed by the Itô formula. The equality of the increments then leads to the equality of the resulting integrands of dt, dW_t^1 and dW_t^2. The equality of the last two integrands implies

$$\begin{aligned} S_t\sqrt{V_t}u'_S &= \gamma_t S_t\sqrt{V_t}c'_S + \beta_t S_t\sqrt{V_t} \\ \theta\sqrt{V_t}u'_V &= \gamma_t\theta\sqrt{V_t}c'_V \end{aligned}$$

so that the stock strategy (β_t) and the option strategy (γ_t) are determined by

$$\begin{aligned} \gamma &= \frac{u'_V}{c'_V} \\ \beta &= u'_S - \gamma c'_S = U'_S - \frac{u'_V c'_S}{c'_V} \end{aligned}$$

Comparing the drifts yields after rearranging

$$\begin{aligned} &\frac{1}{u'_V}\Big(u'_t + S_t\mu u'_S + \kappa(\eta - V_t)u'_V + \frac{1}{2}S_t^2 V_t u''_{SS} + \frac{1}{2}\theta^2 V_t u''_{VV} \\ &+S_t\theta V_t\rho u''_{SV} - ru - (\mu - r)u'_S S_t\Big) \\ =&\frac{1}{c'_V}\Big(c'_t + S_t\mu c'_S + \kappa(\eta - V_t)c'_V + \frac{1}{2}S_t^2 V_t c''_{SS} + \frac{1}{2}\theta^2 V_t c''_{VV} \\ &+S_t\theta V_t\rho c''_{SV} - rc - (\mu - r)c'_S S_t\Big) \end{aligned}$$

where μ is the real world drift of the stock price process.

As this can be derived for any option c (with non vanishing vega) we can conclude that the left hand side is independent of c but only a function of S,

V and t. This function is denoted by λ and called market price of risk. Heston (1993) used a linear form for the market price of risk $\lambda(S_t, V_t, t) = \lambda V_t$. This leads finally the partial differential equation that the value of an option u must satisfy

$$u'_t + Sru'_S + \{\kappa(\eta - V) - \lambda V\}u'_V + \frac{1}{2}V(S^2u''_{SS} + \theta^2u''_{VV} + 2S\theta\rho u''_{SV}) - ru = 0.$$

Depending on the option, boundary and final conditions have to be imposed. The resulting problem can then be solved by numerical techniques like finite difference methods, see e.g. Duffy (2006), or finite elements methods, see e.g. Topper (2005).

Chapter 4

Estimation

In the last chapter, we discussed theoretical properties of option pricing models. The estimation of these option pricing models is analyzed in this chapter on the basis of the example of the Heston model. In section 4.1 we consider the problem to estimate this model from historical stock prices under the real world measure. In section 4.2 we analyze the calibration of the model to option prices and discuss several specifications.

4.1 Estimation from stock prices

The Heston model is an example for a situation where only a part of an evolving system can be observed. In the Heston model, only the stock price can be observed on the market. The variance process is unobservable, it is hidden. In these situations, the Kalman filter is an optimal filter for linear problems. If the problem is nonlinear then there are several approaches for the estimation. We consider here the extended Kalman filter for illustrating the basic idea.

4.1.1 Kalman filter

We consider a general situation where we have a vector of observables $y_t = (y_{1,t}, \ldots, y_{n,t})$ and a vector of state variables $a_t = (a_{1,t}, \ldots, a_{m,t})$. The Kalman filter can be applied if the model can be written as

$$y_t = Z_t(\theta)a_t + d_t(\theta) + \varepsilon_t$$
$$a_{t+1} = T_t(\theta)a_t + c_t(\theta) + R_t(\theta)\eta_t$$

where θ is a vector of parameters, Z_t, T_t and R_t are matrices, d_t and c_t are vectors, ε_t is Gaussian noise with variance H_t and η_t is Gaussian noise with

variance Q_t. $Z_t, T_t, R_t, d_t, c_t, H_t$ and Q_t may depend on y_{t-1} but not on y_t. The first equation is the measurement equation and the second is called system equation.

The estimation is done in three steps – a prediction, an update and a parameter estimation step. Suppose we have at time $t-1$ the current estimates of the state a_{t-1}, the variance P_{t-1} of a_{t-1} and the parameters θ_{t-1}. We present only the algorithm, a mathematical account is given in e.g. Harvey (1989). In the prediction step, a and P are forecasted simply by the unconditional estimates

$$\begin{aligned} a_{t|t-1} &= T_t A_{t-1} + c_t \\ P_{t|t-1} &= T_{t-1} P_{t-1} T'_{t-1} + T_{t-1} Q_{t-1} R'_{t-1} \end{aligned}$$

In the update step, we observe y_t so that the forecast error v_t is

$$v_t = y_t - Z_t a_{t|t-1} - d_t$$

and its variance F_t is

$$F_t = Z_t P_{t|t-1} Z'_t + H_t$$

Hence, the new estimates of a and P are

$$\begin{aligned} a_t &= a_{t|t-1} + P_{t|t-1} Z'_t F_t^{-1} v_t \\ P_t &= P_{t|t-1} - P_{t|t-1} Z'_t F_t^{-1} Z_t P_{t|t-1} \end{aligned}$$

Finally, the parameters can be estimated by maximum likelihood because the errors are assumed to be Gaussian.

4.1.2 Extended Kalman filter

The extended Kalman filter can be applied if the model can be described by

$$\begin{aligned} y_t &= f(a_t, t) + \varepsilon_t \\ a_{t+1} &= g(a_t, t) + R(a_t, t)\eta_t \end{aligned}$$

where f and g are nonlinear. Expanding f and g to first order Taylor series gives the linearised system and measurement equations

$$\begin{aligned} y_t &= f'(\hat{a}_{t|t-1}, t) a_t + f(\hat{a}_{t|t-1}, t) - f'(\hat{a}_{t|t-1}, t) a_{t|t-1} + \varepsilon_t \\ a_{t+1} &= g'(\hat{a}_t, t) a_t + g(\hat{a}_t, t) - g'(\hat{a}_t, t) \hat{a}_t + R(\hat{a}_t, t)\eta_t \end{aligned}$$

Now, the standard Kalman filter of section 4.1.1 can be applied to this linearised equations.

The Heston model (3.4, 3.5) can be estimated in this framework because of the discretization

$$\log S_k = \log S_{k-1} + (\mu - \frac{1}{2}V_{k-1})\Delta t + \sqrt{V_{k-1}}\sqrt{\Delta t}W^1_{k-1}$$
$$V_k = V_{k-1} + (\zeta\eta + \zeta V_{k-1})\Delta t + \theta\sqrt{V_{k-1}}\sqrt{\Delta t}W^2_{k-1}$$

4.2 Calibration to option prices

In the last section, we show how to estimate models from stock prices. In this section, we discuss the estimation in the risk neutral world from option prices. Equity derivative pricing models are calibrated to market data of plain vanilla options by minimization of an error functional. From the economic viewpoint, there are several possibilities to measure the error between the market and the model. These different specifications of the error give rise to different sets of calibrated model parameters and the resulting prices of exotic options vary significantly.

We provide evidence for this calibration risk in a time series of DAX implied volatility surfaces from April 2003 to March 2004. We analyze factors influencing these price differences for exotic options in the Heston and in the Bates model and recommend an error functional. Moreover, we determine the model risk of these two stochastic volatility models for the time series and compare it to calibration risk.

4.2.1 Introduction

Recently, there has been a considerable interest, both from a practical and a theoretical point of view, in the risks involved in option pricing. Schoutens et al. (2004) have analyzed model risk in an empirical study and Cont (2005) has put this risk into a theoretical framework. Another source of risk is hidden in the calibration of models to market data. This calibration risk is also fundamental for the banking industry because it influences significantly the prices of exotic options. Moreover, calibration risk exists even if an appropriate model has been chosen and model risk does not exist anymore.

Calibration risk arises from the different possibilities to measure the error between the observations on the market and the corresponding quantities in the model world. A natural approach to specify this error is to consider the absolute price (AP) differences, see e.g. Schoutens et al. (2004). But the importance of absolute price differences depends on the magnitude of these price. Hence, another useful way for measuring the error are relative price (RP) differences, see e.g. Mikhailov and Nögel (2003). As models are often

judged by their capability to reproduce implied volatility surfaces other measures can be defined in terms of implied volatilities. There are again the two possibilities of absolute implied volatilities (AI) and relative implied volatilities (RI). We consider these four ways to measure the difference between model and market data and explore the implications for the pricing of exotic options.

To this end, we focus on the stochastic volatility model of Heston. In order to analyze the influence of the goodness of fit on calibration risk we consider in addition the Bates model which is an extension of the Heston model with similar qualitative features. These two models are calibrated to the prices of plain vanilla options on the DAX. We use a time series of implied volatility surfaces from April 2003 to March 2004. As exotic options we consider down and out puts, up and out calls and cliquet options for 1, 2 or 3 years to maturity. In this framework we determine the size of calibration risk and analyze factors influencing it.

Besides calibration risk there is also model risk which represents wrong prices because a wrong parametric model has been chosen. We consider the model risk between the Heston and the Bates model and analyze the relation between the two forms of risk in pricing exotic options.

Section 4.2.2 introduces the models and describes their risk neutral dynamics that we use for option pricing. Moreover, this section contains information about the data used for the calibration. Section 4.2.3 describes the calibration method and defines the error functionals analyzed in this work. The goodness of fit is shown by representative surfaces and statistics on the errors. In Section 4.2.4, we present the exotic options that we consider for calibration risk and price these products by simulation. In Section 4.2.5, we analyze the model risk for the two stochastic volatility models under the four error functionals. In the last Section 4.2.6, we summarize the results and draw our conclusions.

4.2.2 Models and Data

In this section, we review the Heston model and describe briefly the Bates model for which we are going to analyze calibration risk. Moreover, we provide some descriptive statistics of the implied volatility surfaces that we use as input data for the calibration.

Heston model

We consider the popular stochastic volatility model of Heston (1993):

$$\frac{dS_t}{S_t} = \mu dt + \sqrt{V_t} dW_t^1$$

where the volatility process is modelled by a square-root process:

$$dV_t = \xi(\eta - V_t)dt + \theta\sqrt{V_t} dW_t^2$$

and W^1 and W^2 are Wiener processes with correlation ρ. A detailed discussion of this model can be found in section 3.1.2.

The variance process (V_t) remains positive if its volatility θ is small enough with respect to the product of the mean reversion speed ξ and the average variance level η:

$$\xi\eta > \frac{\theta^2}{2}. \tag{4.1}$$

The dynamics of the price process are analyzed under a martingale measure under which the characteristic function of $\log(S_t)$ is given by:

$$\begin{aligned}\phi_t^H(z) &= \exp\{\frac{-(z^2+\mathbf{i}z)V_0}{\gamma(z)\coth\frac{\gamma(z)t}{2} + \xi - \mathbf{i}\rho\theta z}\} \\ &\quad \times \frac{\exp\{\frac{\xi\eta t(\xi - \mathbf{i}\rho\theta z)}{\theta^2} + \mathbf{i}ztr + \mathbf{i}z\log(S_0)\}}{(\cosh\frac{\gamma(z)t}{2} + \frac{\xi - \mathbf{i}\rho\theta z}{\gamma(z)}\sinh\frac{\gamma(z)t}{2})^{\frac{2\xi\eta}{\theta^2}}}\end{aligned}$$

where $\gamma(z) \stackrel{\text{def}}{=} \sqrt{\theta^2(z^2+\mathbf{i}z) + (\xi - \mathbf{i}\rho\theta z)^2}$, see e.g. Cont and Tankov (2004a).

Bates model

Bates (1996) extended the Heston model by considering jumps in the stock price process:

$$\begin{aligned}\frac{dS_t}{S_t} &= \mu dt + \sqrt{V_t} dW_t^1 + dZ_t \\ dV_t &= \xi(\eta - V_t)dt + \theta\sqrt{V_t} dW_t^2\end{aligned}$$

where Z is a compound Poisson process with intensity λ and jumps k that have a lognormal distribution:

$$\log(1+k) \sim \mathrm{N}(\log(1+\overline{k}) - \frac{\delta^2}{2}, \delta^2).$$

We analyze the dynamics of this model under a martingale measure under which the characteristic function of $\log(S_t)$ is given by:

$$\begin{aligned}\phi_t^B(z) &= \exp\{t\lambda(e^{-\delta^2 z^2/2+\mathbf{i}\{\log(1+\overline{k})-\frac{1}{2}\delta^2\}z}-1)\} \\ &\quad\times\exp\{\frac{-(z^2+\mathbf{i}z)V_0}{\gamma(z)\coth\frac{\gamma(z)t}{2}+\xi-\mathbf{i}\rho\theta z}\} \\ &\quad\times\frac{\exp\{\frac{\xi\eta t(\xi-\mathbf{i}\rho\theta z)}{\theta^2}+\mathbf{i}zt(r-\overline{k})+\mathbf{i}z\log(S_0)\}}{(\cosh\frac{\gamma(z)t}{2}+\frac{\xi-\mathbf{i}\rho\theta z}{\gamma(z)}\sinh\frac{\gamma(z)t}{2})^{\frac{2\xi\eta}{\theta^2}}}\end{aligned}$$

where $\gamma(z)\stackrel{\text{def}}{=}\sqrt{\theta^2(z^2+\mathbf{i}z)+(\xi-\mathbf{i}\rho\theta z)^2}$, see e.g. Cont and Tankov (2004a).

The Bates model has eight parameters while the Heston model has only five parameters. Because of these three additional parameters the Bates model can better fit observed surface but parameter stability is more difficult to achieve.

Data

Our data consists of EUREX-settlement implied volatilities on the DAX, i.e. if you plug these volatilities into the Black-Scholes formula together with the other corresponding parameters then you get the settlement prices of European options on the DAX. In this context, we approximate the risk free interest rates by the EURIBOR. On each trading day we use the yields corresponding to the maturities of the implied volatility surface. As the DAX is a performance index it is adjusted to dividend payments. Thus, we do not consider dividend payments explicitly.

We analyze the time period from April 2003 to March 2004. Since March 2003 the EUREX trades plain vanilla options with maturities up to 5 years. Until March 2004 it has not changed its range of products. Hence, the data is homogeneous in sense that the implied volatility surfaces are derived from similar products.

From this time period we analyze the surfaces from all the Wednesdays when trading has taken place. We restrict ourselves to these days because of the computationally intense Monte Carlo simulations for the pricing. Thus, we consider 51 implied volatility surfaces. We exclude observations that are deep out of the money because of illiquidity of these products. More precisely, we consider only options with moneyness $K/S_0\in[0.75,1.35]$ for small times to maturity $T\leq 1$. As we analyze exotic options that expire in 1, 2 or 3 years we exclude also plain vanillas with time to maturity less than 3 months.

Some information about the resulting implied volatility surfaces are summarized in table 4.1 . The surfaces contain on average 140 prices and nine times to maturity with a mean moneyness range of 65%.

	mean number of maturities	mean number of observations	mean moneyness range
short maturities ($0.25 \leq T < 1.0$)	3.06	64	0.553
long maturities ($1.0 \leq T$)	5.98	76	0.699
total	9.04	140	0.649

Table 4.1: Description of the implied volatility surfaces.

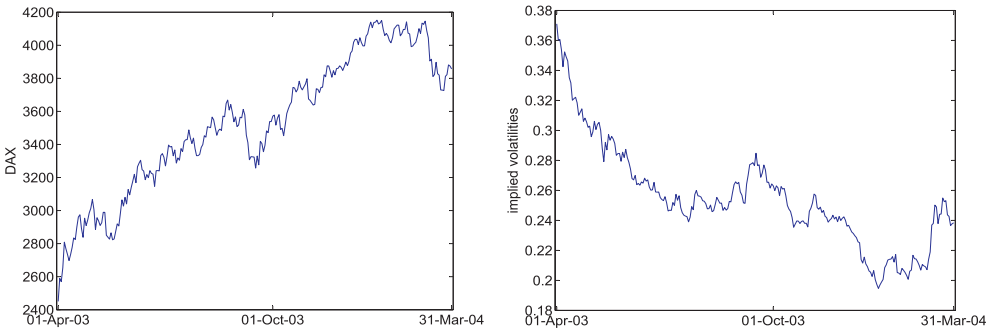

Figure 4.1: DAX and ATM implied volatility with 1 year to maturity on the trading days from 01 April 2003 to 31 March 2004.

The values of the underlying in the sample period are shown in figure 4.1. This figure also shows the (interpolated) at the money implied volatilities for 1 year to maturity. The market value of the DAX went up in this period and accordingly the implied volatilities went down as figure 4.1 shows.

4.2.3 Calibration

In this section, we specify the calibration routine and describe the four error functionals. The calibration results illustrate how good the plain vanilla prices can be replicated by the Heston and the Bates model.

Calibration method

Carr and Madan (1999) found a representation of the price of a European call option by one integral for a whole class of option pricing models. Their

method that is applicable to the Heston and the Bates model is based on the characteristic function of the log stock price under the risk neutral measure.

Carr and Madan showed that the price $C(K,T)$ of a European call option with strike K and maturity T is given by

$$C(K,T) = \frac{\exp\{-\alpha \ln(K)\}}{\pi} \int_0^{+\infty} \exp\{-\mathbf{i}v \ln(K)\} \psi_T(v) dv$$

for a (suitable) damping factor $\alpha > 0$. The function ψ_T is given by

$$\psi_T(v) = \frac{\exp(-rT)\phi_T\{v - (\alpha+1)\mathbf{i}\}}{\alpha^2 + \alpha - v^2 + \mathbf{i}(2\alpha+1)v}$$

where ϕ_T is the characteristic function of $\log(S_T)$, see Section 4.2.2.

For the difference between market and model we consider the following four objective functions based on the root weighted square error:

$$\begin{aligned}
\text{AP} &\stackrel{\text{def}}{=} \sqrt{\sum_{i=1}^{n} w_i (P_i^{mod} - P_i^{mar})^2} \\
\text{RP} &\stackrel{\text{def}}{=} \sqrt{\sum_{i=1}^{n} w_i (\frac{P_i^{mod} - P_i^{mar}}{P_i^{mar}})^2} \\
\text{AI} &\stackrel{\text{def}}{=} \sqrt{\sum_{i=1}^{n} w_i (IV_i^{mod} - IV_i^{mar})^2} \\
\text{RI} &\stackrel{\text{def}}{=} \sqrt{\sum_{i=1}^{n} w_i (\frac{IV_i^{mod} - IV_i^{mar}}{IV_i^{mar}})^2}
\end{aligned}$$

where *mod* refers to a model quantity and *mar* to a quantity observed on the market, P to a price and IV to an implied volatility. The index i runs over all n_t observations of the surface on day t. The weights w_i are non negative with $\sum_i w_i = 1$. Hence, the objective functions can be interpreted as mean average errors. While the error functionals AP and RP measure the differences between option prices, the other two error measures focus on Black-Scholes implied volatilities because these quantities are normally used in reality for price quotations. The model implied volatilities IV^{mod} are computed from the prices of European options in the Heston and in the Bates model by numerical inversion of the Black-Scholes formula.

We choose the weights in such a way that on each day all maturities have the same influence on the objective function. In order to make different surfaces comparable each maturity gets the weight $1/n_{mat}$ where n_{mat} denotes

the number of maturities in this surface. Moreover, we assign the same weight to all points of the same maturity. This leads to the weights

$$w_i \stackrel{\text{def}}{=} \frac{1}{n_{mat} n^i_{str}}$$

where n^i_{str} denotes the number of strikes with the same maturity as observation i. This weighting leads asymptotically to a uniform density on each maturity.

Given these weights the average time to maturity of an implied volatility surface can be measured by a modified duration:

$$\sum_{i=1}^{n} \frac{\tau_i w_i}{\sum_{i=1}^{n} w_i}$$

where τ_i is the time to maturity of the option i. The mean duration of the 51 surfaces is 2.02 and the minimal (maximal) is 1.70 (2.30). Thus, the point of balance for the maturities lies around 2 for our time series of surfaces. As we analyze exotic options with 1, 2 or 3 years time to maturity this point of balance confirms a correct weighting for our purposes.

We consider only out of the money prices. Thus, we use call prices for strikes higher than the spot and put prices for strikes below the spot. This approach ensures to compare only prices of similar magnitude. It has no impact on the errors based on implied volatilities. Because of the put call parity the use of OTM options has nor an impact on the absolute price error (AP). But the relative prices are weighted in such a way that the observations around the spot receive less weight. Hence, only the relative price error (RP) is influenced by this choice of prices.

In order to estimate the model parameters we apply a stochastic global optimization routine and minimize the objective functions with respect to the model parameters. In addition to some natural constraints on the range of the parameters we have taken into account inequality (4.1) that ensures the positivity of the variance process.

Calibration results

We consider 51 implied volatility surfaces between from April 2003 and March 2004. Each of these is calibrated with respect to the four error functions described in Section 4.2.3. These calibrations are done for the Heston and the Bates model.

The resulting errors of these 408 calibrations have been summarized in table 4.2 for the Heston model and in table 4.3 for the Bates model. Descriptive statistics on the calibrated parameters are given in table 4.4 for

mean objective fct.	AP	RP $[E^{-2}]$	AI $[E^{-2}]$	RI $[E^{-2}]$
AP	7.3	9.7	0.81	3.1
	(2.2)	(4.8)	(0.25)	(1.2)
RP	11.	6.1	0.74	2.9
	(4.6)	(2.5)	(0.22)	(1.0)
AI	9.4	7.3	0.68	2.6
	(3.2)	(3.2)	(0.20)	(0.9)
RI	8.8	7.0	0.70	2.5
	(2.9)	(3.0)	(0.21)	(0.9)

Table 4.2: Mean calibration errors in the Heston model for 51 days. (AP=absolute price differences, RP=relative price differences, AI=absolute implied volatility differences, RI=relative implied volatility differences)

the Heston model and in table 4.5 for the Bates model. Figure 4.2 shows the fit of the implied volatility surface in the Heston model on a day that is representative for the AI error.

Table 4.2 reports in each line the means of the four errors when the objective function given in the left column is minimized. In the Heston model, we get a mean absolute price error of 7.3 and a mean relative price error of 9.7% when we calibrate with respect to AP. Using the RP error functional we get the opposite result with a mean absolute price error of 11 and a mean relative price error of 6.1%. The errors based on implied volatilities are smaller for the RP objective function than for the AP objective function. The results for the AI and RI objective functionals differ only slightly: the mean absolute implied volatility error is about 0.68% and the mean relative implied volatility error is about 2.5%. Moreover, the price errors for these objective functions lie between the price errors of the other two objective functions. The calibration with respect to RI gives the best overall fit because it has the smallest RI error and the second best errors for the rest. The meaning of these numbers is illustrated by figure 4.2 which shows an implied volatility fit that is representative for an AI error of 0.68%. In order to make the AP errors comparable for different days (with different values of the spot) we have computed the mean of AP/DAX as $0.21\%, 0.34\%, 0.27\%, 0.25\%$ for the four error functionals.

The calibrated parameters which are described by table 4.4 form two groups because the parameters for the RP, AI and RI calibration are quite similar. The initial variance V_0 and the average variance level η are both about 0.07 for all objective functionals. For the AP calibration we get a

mean objective fct.	AP	RP $[E^{-2}]$	AI $[E^{-2}]$	RI $[E^{-2}]$
AP	7.0	13.	0.76	2.8
	(2.2)	(10)	(0.23)	(0.9)
RP	12.	5.1	0.67	2.6
	(4.8)	(1.9)	(0.20)	(0.9)
AI	8.9	6.4	0.60	2.3
	(3.2)	(2.6)	(0.18)	(0.8)
RI	8.7	6.2	0.62	2.2
	(3.2)	(2.8)	(0.20)	(0.8)

Table 4.3: Mean calibration errors in the Bates model for 51 days. (AP=absolute price differences, RP=relative price differences, AI=absolute implied volatility differences, RI=relative implied volatility differences)

	ξ	η	θ	ρ	V_0
AP	0.87	0.07	0.34	-0.82	0.07
	(0.48)	(0.02)	(0.08)	(0.08)	(0.02)
RP	1.38	0.07	0.44	-0.74	0.08
	(0.35)	(0.02)	(0.06)	(0.03)	(0.02)
AI	1.32	0.07	0.43	-0.77	0.08
	(0.40)	(0.02)	(0.06)	(0.04)	(0.02)
RI	1.20	0.07	0.41	-0.75	0.08
	(0.35)	(0.02)	(0.06)	(0.05)	(0.02)

Table 4.4: Mean parameters (std.) in the Heston model for 51 days.

	ξ	η	θ	ρ	V_0	λ	$\overline{k}$	δ
AP	0.92	0.07	0.33	-0.94	0.07	0.33	0.07	0.08
	(0.50)	(0.02)	(0.08)	(0.07)	(0.02)	(0.21)	(0.03)	(0.06)
RP	1.56	0.07	0.45	-0.89	0.08	0.54	0.05	0.08
	(0.47)	(0.02)	(0.07)	(0.07)	(0.02)	(0.23)	(0.03)	(0.06)
AI	1.43	0.07	0.43	-0.95	0.07	0.50	0.06	0.09
	(0.44)	(0.02)	(0.06)	(0.06)	(0.02)	(0.22)	(0.03)	(0.04)
RI	1.36	0.07	0.41	-0.93	0.07	0.52	0.05	0.08
	(0.44)	(0.02)	(0.07)	(0.09)	(0.02)	(0.26)	(0.04)	(0.08)

Table 4.5: Mean parameters (std.) in the Bates model for 51 days.

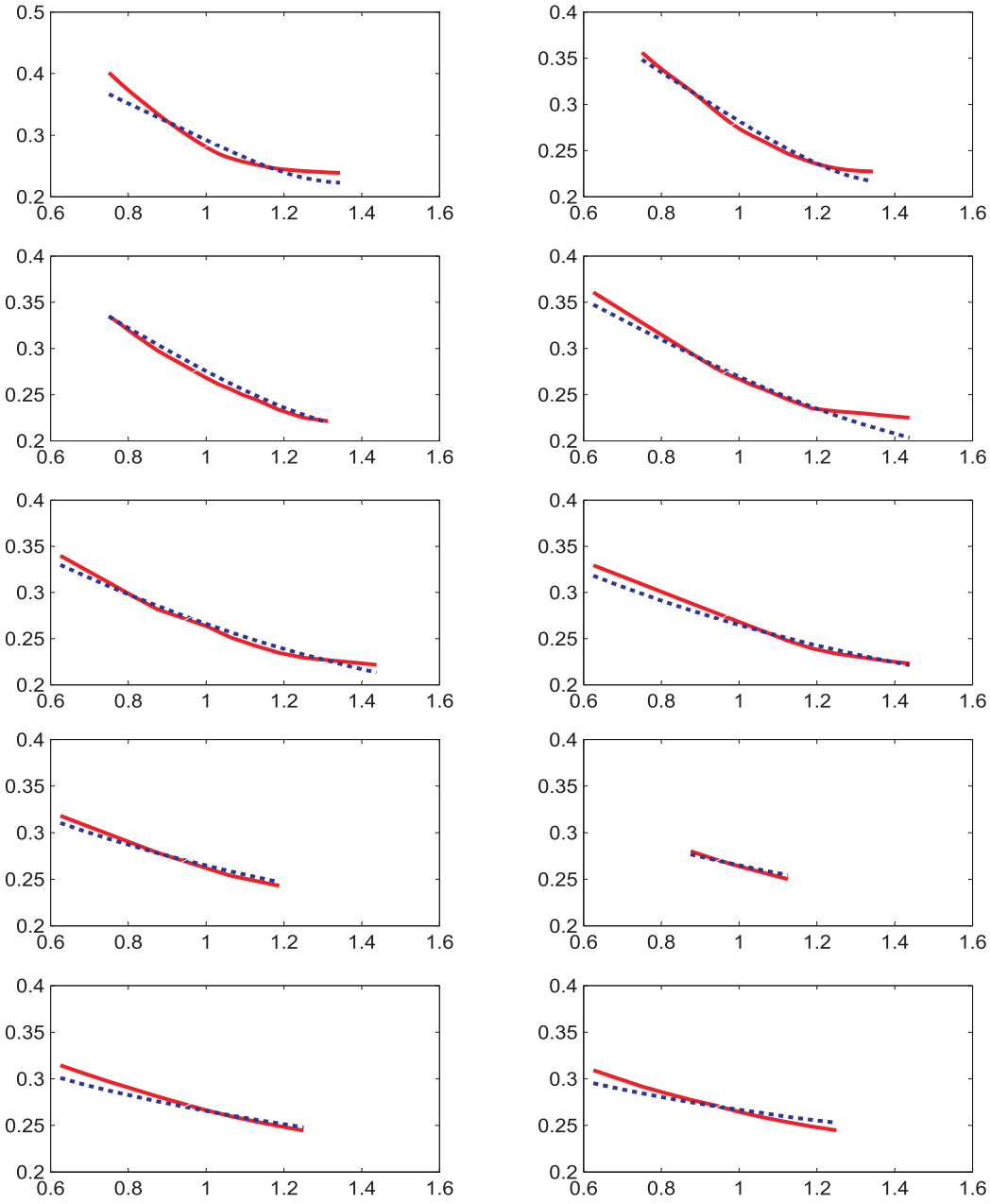

Figure 4.2: Implied volatilities in the Heston model for the maturities 0.26, 0.52, 0.78, 1.04, 1.56, 2.08, 2.60, 3.12, 3.64, 4.70 (left to right, top to bottom) for AI parameters on 25/06/2003. Solid: model, dotted: market. X-axis: moneyness.

reversion speed $\xi = 0.9$, a volatility of variance of $\theta = 0.34$ and a correlation $\rho = -0.82$. The other calibrations lead to similar parameters with a reversion speed $\xi = 1.3$, a volatility of volatility of $\theta = 0.44$ and a correlation $\rho = -0.75$. As the correlations are significantly below -1 the calibrated Heston models have really two stochastic factors.

The Bates model exhibits similar qualitative results as the Heston model: The AP and the RP calibrations differ clearly while the AI and the RI calibrations lead to similar results. The Bates model can be regarded as an extension of the Heston model. The additional three parameters for the jumps in the spot process lead for all errors functionals to better calibration results: The AP error is reduced (on average) by 4%, the RP error by 16%, the AI and the RI error both by 12%.

The calibrated parameters of the Bates model are given in table 4.5. As in the Heston model they form two groups with the AP calibration on the one hand and the RP, AI and RI calibrations on the other hand. The parameters ξ, η, θ and V_0 are similar to the calibrations for the Heston model. Only the correlation ρ rises to a level of -0.93 for all objective functions. Hence, this criterion for distinguishing between the two groups disappears. It is replaced by the expected number of jumps per year: For the AP calibration we expect (on average) a jump every three years while we expect every two years a jump for the other calibrations. It is interesting that all calibrations lead to a mean jump up of about +8% for the returns. The expected jumps upwards correspond to the market going up as shown in figure 4.1.

Schoutens et al. (2004) found that the Heston and the Bates option model can both be calibrated well to the EuroStoxx50. In summarizing the results of this section we can say that DAX implied volatility surfaces can be replicated by these models for different error functionals as good as the EuroStoxx50 data in Schoutens et al. (2004). As in that work, we find that the Bates model gives only slightly better fits for the AP calibration. In addition we have shown that it leads to a considerable improvement in the fit for the other objective functions.

4.2.4 Exotic Options

In this section, we analyze the price differences of exotic options for calibrations with respect to different error measures. We consider barrier and cliquet options. The prices of these products are calculated by Monte Carlo simulations using Euler discretizations.

	Heston			Bates		
	$T=1$	$T=2$	$T=3$	$T=1$	$T=2$	$T=3$
up and out calls	0.002	0.001	0.001	0.002	0.001	0.001
down and out puts	0.002	0.001	0.001	0.002	0.001	0.001
cliquet options	0.001	0.001	0.001	0.001	0.001	0.001

Table 4.6: Maximal ratio of standard error and price in Monte Carlo simulations. (Maximum over all time points and all objective functions)

Simulation

We price all exotic options by Monte Carlo simulations. To this end, we use for each derivate 1000000 paths generated by Euler discretization, see e.g. Glasserman (2004). As we take into account the positivity constraint (4.1) the square root process for the variance can be simulated by truncation at zero

$$V_{t_{i+1}} = \left(V_{t_i} + \xi(\eta - V_{t_i})\Delta_t + \theta\sqrt{V_{t_i}}\sqrt{\Delta_t}Z_i \right)^+$$

where Δ_t is the time step and Z_i are independent standard normal variables. This simple scheme leads to an acceptable small bias when the positivity constraint is fulfilled. For each exotic option we consider three maturities: 1 year, 2 years and 3 years. We analyze three exotic options: up and out calls, down and out puts and cliquet options. These products are described in the following sections where remaining parameters are also specified.

The payoffs of barrier options depend only on whether the underlying price process exceeded the barrier in some time interval. Hence, the value of barrier options depends on the minimum or maximum of the underlying price process. We approximate such continuous extrema by discrete extrema using one observation for each trading day. We use 252 time steps to simulate a process for a year assuming 252 trading days a year.

The calibration results are presented in following sections together with a discussion of the options. The accuracy of the Monte Carlo results is given by the relative standard error in table 4.6. This table confirms that the estimators have sufficiently small variance after 1000000 paths compared to the price differences that we observe in tables 4.7 to 4.9.

Barrier Options

We consider two types of barrier options: up and out calls and down and out puts. These options are quite popular on the market. Down and out puts are sold e.g. together with zero-strike calls as bonus certificates. These structured products are actively traded in Germany and also in many other markets.

Up and out call options

The prices of up and out calls with strike K, barrier B and maturity T on an underlying (S_t) are given by

$$\exp(-rT)\,\mathrm{E}[\mathbf{1}_{\{M_T<B\}}(S_T-K)^+]$$

where

$$M_T \stackrel{\text{def}}{=} \max_{0\le t\le T} S_t.$$

We choose as strike K and barrier B

$$K = (1-0.1T)S_0$$
$$B = (1+0.2T)S_0$$

where T denotes time to maturity. Up and out calls with such strikes and barriers are widely traded on the German market.

Up and out call options have the payoff profile of European call options if the underlying has not exceeded the barrier. Otherwise their payoff is zero. Thus, up and out calls are path dependent exotic options.

We want to analyze the difference between the prices of the exotic options when the underlying model has been calibrated with respect to different error measures. To this end, we have calibrated the Heston and the Bates model to implied volatility or price data on each day with respect to the four error functionals introduced in Section 4.2.3. Hence, we have four time series of calibrated model parameters that are described in Section 4.2.3. By Monte Carlo simulations we calculate on each day the prices of up and out calls for the four sets of model parameters. In this way we get four time series of up and out call prices corresponding to the four error measures for the Heston model and four corresponding time series of prices for the Bates model. We are interested how the prices of exotic options differ when the four different error functionals are used. If the prices of up and out calls that are computed

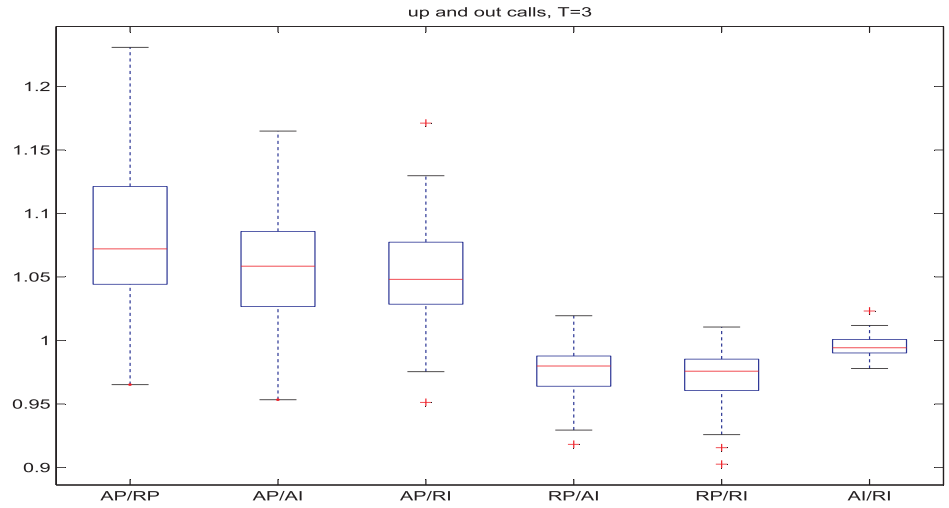

Figure 4.3: Relative prices of the up and out calls in the Heston model for 3 years to maturity.

from the AP parameters are denoted by P_{AP}^{UOC} and the corresponding prices from the RP parameters by P_{RP}^{UOC} then we measure the difference between these prices by the ratio $P_{AP}^{UOC}/P_{RP}^{UOC}$. The other five price differences are measured by corresponding price ratios. Hence, we observe on each day six price ratios that describe the differences of the prices of exotic options resulting from different error measures.

The six possible time series of price ratios are summarized in figure 4.3 for up and out calls with 3 years to maturity in the Heston model. In the boxplots the central line gives the median and the box contains 50% of the observations. Hence, the AP prices lie on average about 6% over the other prices and the AP prices are in 75% of the 51 days at least 4% higher than the other prices. The RP prices are about 2% below the AI or RI prices which are very similar to each other.

We analyze the influence of time to maturity on these price differences by considering also 1 year and 2 years to maturity (and by adjusting the barrier and the strike appropriately). The medians of the price ratios are presented in table 4.7 for all three times to maturity. This table shows that the price differences become smaller for shorter times to maturity for the AP prices. The other price differences remain almost constant. For 1 year to maturity the price differences are about $2\% - 3\%$ and the AP prices are lower than the other prices. For 2 years to maturity the AP prices are again higher than the other prices.

		AP/RP	AP/AI	AP/RI	RP/AI	RP/RI	AI/RI
Heston	$T = 1$	0.986	0.968	0.967	0.984	0.984	0.999
	$T = 2$	1.051	1.024	1.022	0.979	0.978	0.998
	$T = 3$	1.072	1.059	1.048	0.980	0.976	0.994
Bates	$T = 1$	0.988	0.985	1.002	1.002	1.006	1.012
	$T = 2$	1.070	1.083	1.104	0.970	0.986	1.018
	$T = 3$	1.106	1.123	1.129	0.972	0.975	1.013

Table 4.7: Median of price ratios of up and out calls.

In order to analyze the influence of the goodness of fit on the price differences we consider also the Bates model. The boxplots of the price ratios in this model are given in figure 4.4 for 3 years to maturity. Compared to the Heston boxplots the boxes are longer in the Bates model. Thus there is more variation between the prices for different error functionals. Moreover the median differences between the AP prices and the other prices are bigger than in the Heston model - especially for AP/AI and AP/RI. The differences between RP, AI and RI are similar to those in the Heston model. The corresponding results for 1 year and 2 years to maturity are presented in table 4.7. Qualitatively the situation is similar to the Heston model: For shorter times to maturity the price differences decrease - especially for AP prices.

Thus, the price differences in the Heston and in the Bates model are similar between the RP, AI and RI prices while the AP price differences are bigger in the Bates model. Moreover, the variation of the price differences is higher in the Bates model.

Down and out put options

The prices of the down and out puts with strike K, barrier B and maturity T on an underlying (S_t) are given by

$$\exp(-rT)\,\mathrm{E}[\mathbf{1}_{\{m_T > B\}}(K - S_T)^+]$$

where

$$m_T \stackrel{\text{def}}{=} \min_{0 \le t \le T} S_t.$$

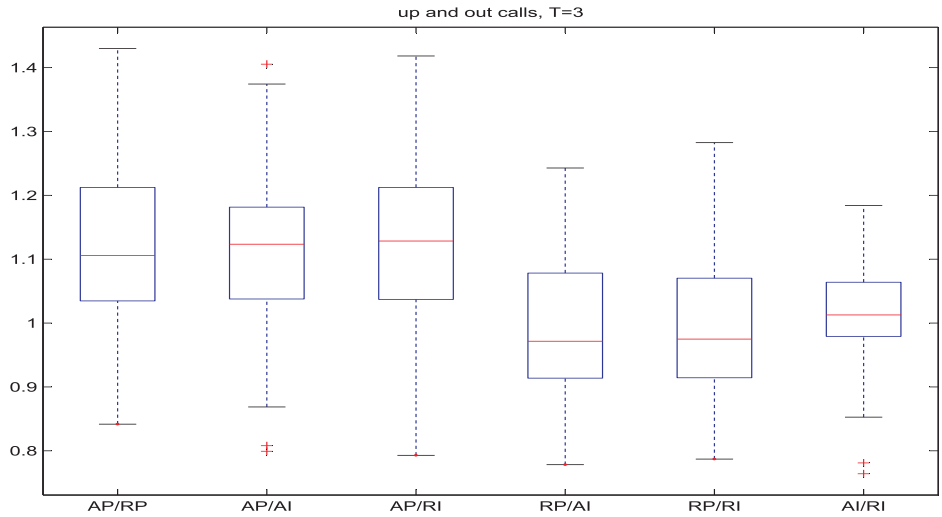

Figure 4.4: Relative prices of the up and out calls in the Bates model for 3 years to maturity.

For our analysis, we use the strike K and the barrier B

$$K = (1 + 0.1T)S_0$$
$$B = (1 - 0.2T)S_0$$

where T denotes time to maturity. The strikes and barriers are set analogously the up and out calls. Such down and out puts are often part of bonus certificates. A bonus certificate is a structured product incorporating a zero-strike call and a down and out put option. The payoff structure at expiry can be described as follows: If the value of the underlying at expiry is above the strike then the investors receive the underlying. If the price of the underlying is below the barrier then the investor bears the full loss of the underlying. Otherwise the investors the underlying and in addition a bonus if the barrier has not been exceeded before. Because of this payoff profile, this product is constructed mainly for sidewards markets.

Down and out put options have the payoff profile of European put options if the underlying has been above the barrier during the life time of the option. Otherwise their payoff is zero.

As described above, we calculate on each day the prices of the down and out puts for the four parameter sets. The resulting six time series of price ratios are shown in the figure 4.5 for 3 years to maturity in the Heston model. The AP prices are (in the mean) about 3.5% smaller than the other prices

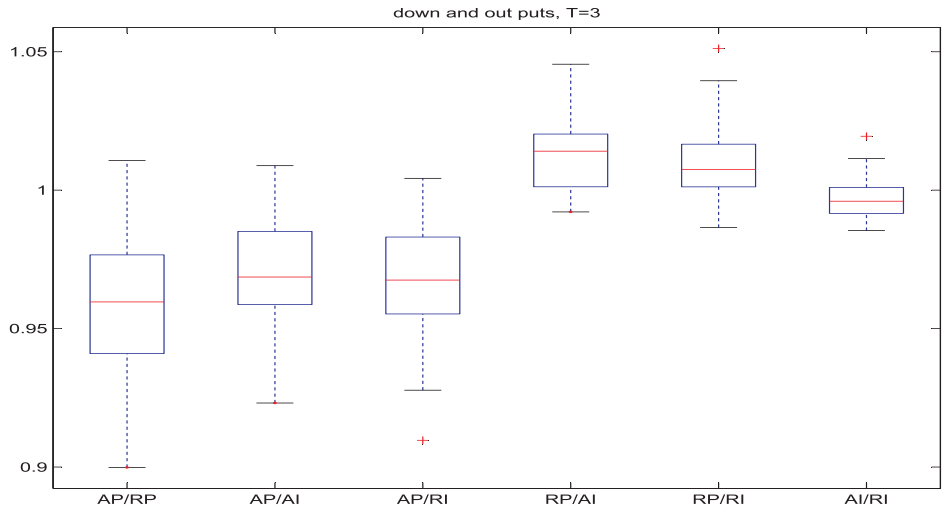

Figure 4.5: Relative prices of the down and out puts in the Heston model for 3 years to maturity.

and 75% of the AP prices are at least 2% smaller than the other prices. The RP prices lie above the prices from the calibrations to implied volatilities. These AI and RI prices are quite similar so that we can identify again the two groups that we have already observed for the up and out calls.

Compared to the up and out calls the price differences are smaller for the down and out puts. This can be seen also from table 4.8 that reports the median of the price ratios for $1, 2$ and 3 years to maturity. This table shows that the price differences change for increasing time to maturity: For 1 year to maturity the AP prices lie above the other prices but with increasing time to maturity the AP prices become relatively smaller. The RP and AI prices remain on a similar level for all times to maturity and the RI prices tend to this level for longer times to maturity.

The situation in the Bates model that gives better fits to the plain vanilla data is given by figure 4.6 and table 4.8. The AP prices lie about 7% below the other prices. Thus this difference is bigger than in the Heston model. The other price ratio lie still on average on the same level but the their variance has grown compared to the Heston model.

The situation for the barrier options can be summarized as follows: The AP prices differ significantly from the other prices for both types of barrier options. While the AP prices are higher for up and out calls they are lower for down and out puts relatively to the other prices. In this sense the situation is

		AP/RP	AP/AI	AP/RI	RP/AI	RP/RI	AI/RI
Heston	$T=1$	1.025	1.031	1.005	1.007	0.980	0.977
	$T=2$	0.983	0.994	0.984	1.011	0.997	0.986
	$T=3$	0.960	0.969	0.968	1.014	1.008	0.996
Bates	$T=1$	1.021	1.012	1.019	1.004	1.006	0.998
	$T=2$	0.968	0.975	0.966	1.031	1.022	0.990
	$T=3$	0.922	0.935	0.931	1.026	1.022	0.995

Table 4.8: Median of price ratios of down and out puts.

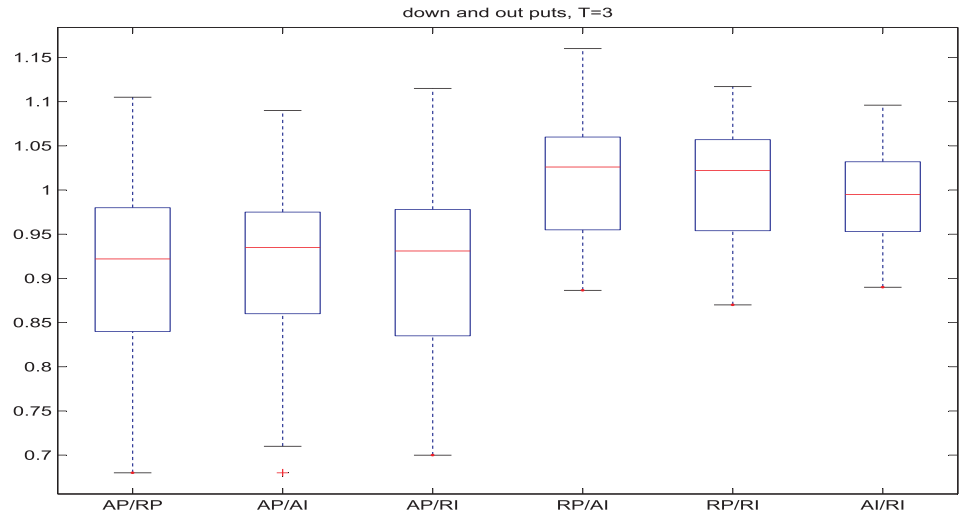

Figure 4.6: Relative prices of the down and out puts in the Bates model for 3 years to maturity.

symmetrical. The differences become bigger for longer times to maturity and the better fit of the Bates model does not lead to smaller price differences.

Cliquet Options

We consider cliquet options with prices

$$\exp(-rT)\,\mathrm{E}[H]$$

where the payoff H is given by

$$H \stackrel{\text{def}}{=} \min(c_g, \max[f_g, \sum_{i=1}^{N} \min\{c_l^i, \max(f_l^i, \frac{S_{t_i} - S_{t_{i-1}}}{S_{t_{i-1}}})\}]).$$

Here c_g (f_g) is a global cap (floor) and c_g^i (f_g^i) is a local cap (floor) for the period $[t_{i-1}, t_i]$.

We consider three periods with $t_i = \frac{T}{3} i$ $(i = 0, \ldots, 3)$ and the caps and floors are given by

$$\begin{aligned} c_g &= \infty \\ f_g &= 0 \\ c_l^i &= 0.08,\ i = 1, 2, 3 \\ f_l^i &= -0.08,\ i = 1, 2, 3 \end{aligned}$$

While barrier options are simple exotic options, cliquet options are more difficult because of their forward structure. Moreover, there exist many different types corresponding to the parameters. Hence, our specification cannot give a representative picture of all the traded cliquets. But these caps and floors are typical because the option holder cannot loose money and the returns are bounded above only by the local return bounds.

Cliquet options pay out basically the sum of the returns $R_i \stackrel{\text{def}}{=} (S_{t_i} - S_{t_{i-1}})/S_{t_{i-1}}$. In order to reduce risk local and global floors f are introduced for the returns R. In the same way the returns are bounded from above by local and global caps c.

The distributions of the six time series of price ratios for cliquet options are described in figure 4.7 for 3 years to maturity in the Heston model. The differences are smaller than in the case of the barrier options. The AP prices lie above the other prices but the difference is significant only for the AP and RP prices. The differences between the other prices is also small. Thus, we cannot recognize directly from this figure the two groups that we identified for the barrier options.

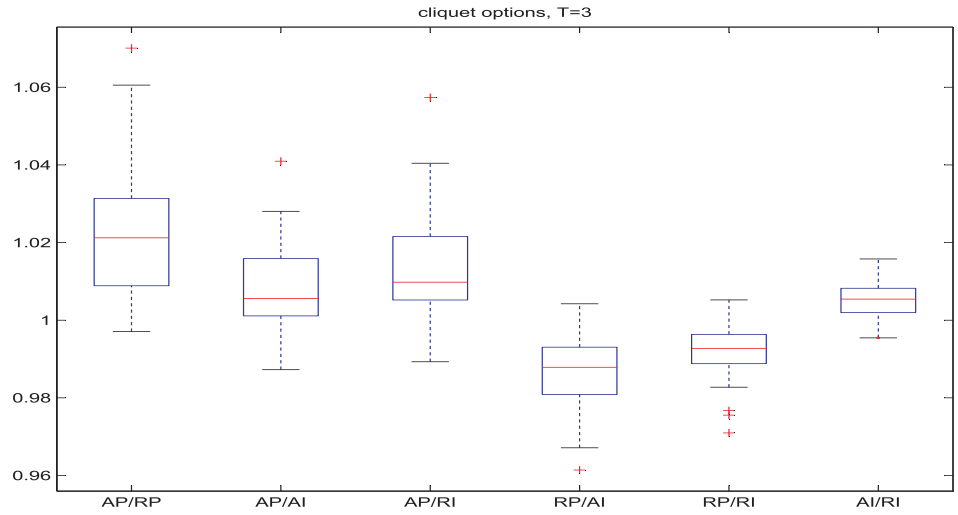

Figure 4.7: Relative prices of the cliquet options in the Heston model for 3 years to maturity.

Table 4.9 that reports the median price differences for 1, 2 and 3 years to maturity gives some insight into this situation: The AP prices are about 2% smaller than the other prices for 1 year to maturity. With increasing time to maturity the AP prices grow relatively and are about 1.5% higher than the other prices for 3 years to maturity. As table 4.9 confirms the other prices remain relatively constant for different times to maturity. Thus there are again the two groups that we have identified for the barrier options: The changing AP prices on the one hand and the constant other prices on the other hand.

		AP/RP	AP/AI	AP/RI	RP/AI	RP/RI	AI/RI
Heston	$T=1$	0.983	0.976	0.989	0.993	1.006	1.013
	$T=2$	1.002	0.991	1.000	0.989	0.998	1.010
	$T=3$	1.022	1.008	1.014	0.987	0.992	1.005
Bates	$T=1$	0.917	0.899	0.917	0.987	1.005	1.024
	$T=2$	0.931	0.903	0.923	0.980	0.999	1.029
	$T=3$	0.946	0.912	0.933	0.976	0.995	1.029

Table 4.9: Median of price ratios of cliquet options.

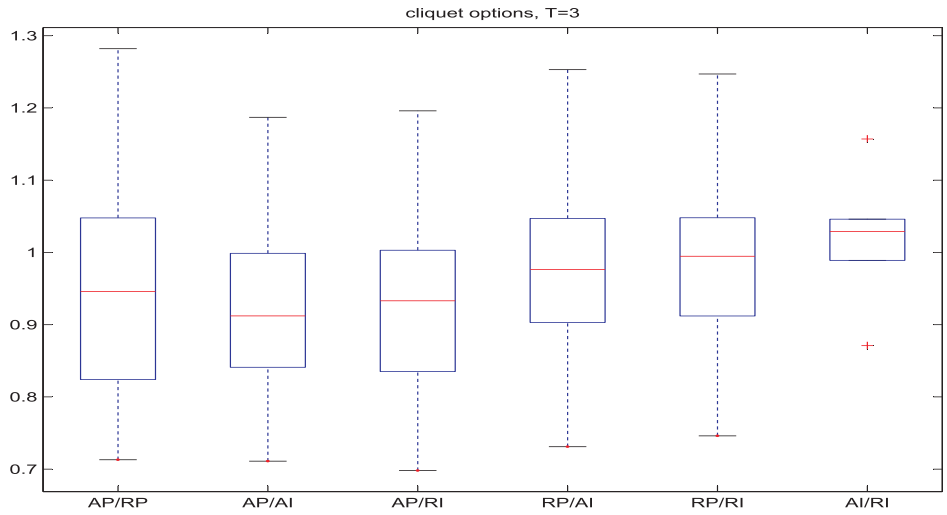

Figure 4.8: Relative prices of the cliquet options in the Bates model for 3 years to maturity.

The relative prices of the cliquet options in the Bates model are presented in figure 4.8 for 3 years to maturity. Here we see that the AP prices are about 7% smaller than the other prices. The RP prices lie about 2% under the AI prices that are 3% higher than the RI prices. The RP and RI prices are similar. Thus, there are quite big differences for the cliquet options in the Bates model. Moreover, the variance is larger relative to the Heston model. Table 4.9 describes the situation of different times to maturity and shows that the AP prices grow relatively with increasing time to maturity while the other prices remain relatively constant for different times to maturity.

Comparing the results for the two barrier options and the cliquets we see in all cases two groups, the AP prices and the other prices. The AP prices differ a lot from the other prices and in addition change relatively for different times to maturity. Moreover, the variance of the price ratio with AP prices is bigger in general than for the other price ratios. The other group of RP, AI and RI prices shows similar prices and small variances. The Bates model that gives better fits has higher price differences (with higher variances).

4.2.5 Model risk

In the last section, we have described the price differences that result from the calibration with respect to the four error functionals. In this section we

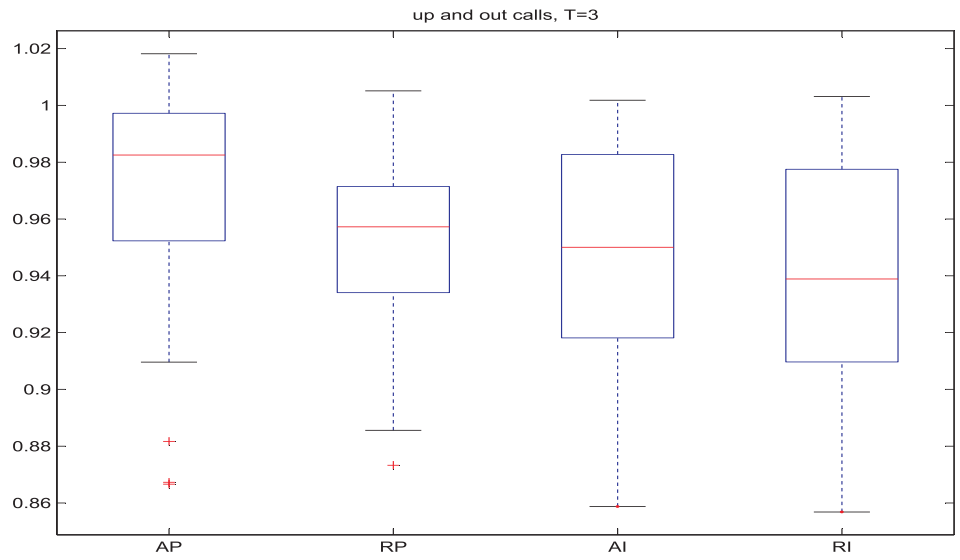

Figure 4.9: Bates prices over Heston prices for up and out calls with 3 years to maturity on 51 days.

consider model risk, consider its relation to calibration risk and compare our results with the findings of Schoutens et al. (2004). Model risk is generally understood as the risk of "wrong" prices because an inappropriate parametric model has been chosen for the stochastic process of the underlying.

In order to analyze this model risk for the two stochastic volatility models, we consider the ratios of the prices of the exotic options in the Bates model and the corresponding prices in the Heston model. The distribution of these ratios for up and out calls with 3 years to maturity is described by the figure 4.9. The prices in the Bates model lie below the prices in the Heston model for all four error functionals: The difference varies between 2% for the AP prices and 6% for the RI prices. Thus model risk is not independent of the calibration method, i.e. calibration risk. The results for smaller times to maturity are given in table 4.10. The table suggests that model risk does not change significantly for different times to maturity.

The model risk of down and out puts is shown in figure 4.10 for 3 years to maturity. The prices in the Bates model lie below the prices in the Heston model for all error functionals. Compared to the up and out calls the model risk is bigger for the down and out puts: It varies between 9% for AP prices and 14% for RI prices. But again we observe the highest difference for RI prices and the smallest for AP prices. Moreover, the variance is bigger than for the up and out calls. Table 4.10 that gives the results for smaller times

		AP	RP	AI	RI
up and out calls	$T = 1$	0.973	0.953	0.944	0.941
	$T = 2$	0.980	0.954	0.953	0.940
	$T = 3$	0.983	0.957	0.950	0.939
down and out puts	$T = 1$	0.933	0.892	0.877	0.878
	$T = 2$	0.918	0.883	0.872	0.860
	$T = 3$	0.916	0.881	0.873	0.860
cliquets	$T = 1$	1.057	1.100	1.109	1.119
	$T = 2$	1.076	1.128	1.130	1.144
	$T = 3$	1.086	1.138	1.140	1.162

Table 4.10: Median of Bates prices over Heston prices.

to maturity suggests that the model risk becomes smaller for shorter times to maturity.

Finally, we consider the model risk of cliquet options in figure 4.11. For these options the Bates prices lie above the corresponding Heston prices for all calibration methods. The smallest price difference that appears for the AP prices is about 8% while the biggest difference of 16% have the RI prices. Table 4.10 shows again smaller price differences for shorter times to maturity.

The model risk between the Heston and the Bates model can be described for barrier and cliquet options as follows: Model risk measured by the price differences in the two models increases for longer times to maturity. Moreover, it is ordered with respect to the calibration method. The calibration with respect to implied volatilities leads to bigger price differences as calibration with respect to prices. The model risk is smallest for AP calibration and bigger for RP calibration. It is even bigger for the AI calibration and the price differences are the biggest for RI calibrations. This emphasizes once more the importance of the implied volatility surfaces and their calibration. Moreover, model risk differs across option types.

Schoutens et al. (2004) consider up and out calls (with strike equal to spot) and cliquet options with 3 years to maturity. For a barrier 50% above the spot, they find a model risk for the up and out calls of about 14%. For the cliquet options Schoutens et al. do not find a significant model risk. These results do not correspond in every respect to our AP results. There may be several reasons for these different results: While we look at a time series of 51 implied volatility surfaces they focus only on one day. Moreover, they have analyzed the EuroStoxx50 and we use DAX data.

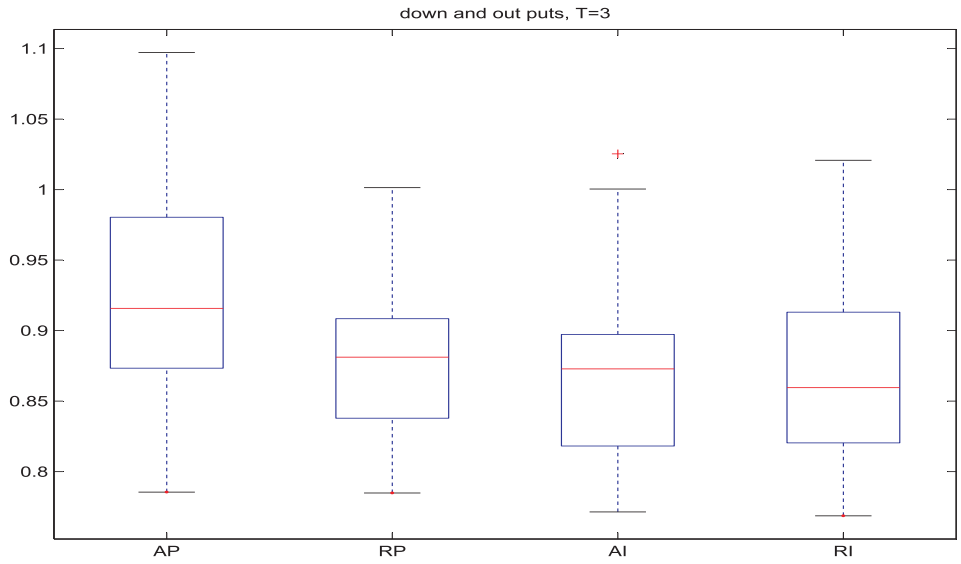

Figure 4.10: Bates prices over Heston prices for down and out puts with 3 years to maturity on 51 days.

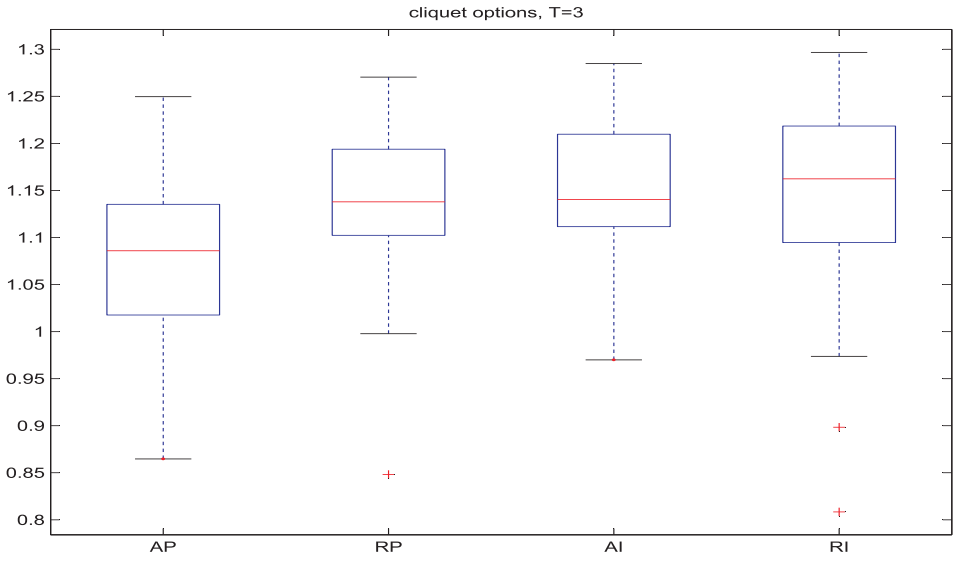

Figure 4.11: Bates prices over Heston prices for cliquet options with 3 years to maturity on 51 days.

4.2.6 Conclusion

We have looked at the stochastic volatility model of Heston and analyzed different calibration methods and their impact on the pricing of exotic options. Our analysis has been carried out for a time series of DAX implied volatility surfaces from April 2003 to March 2004.

We have shown that different ways to measure the error between the model and the market in the calibration routine lead to significant price differences of exotic options. We have considered the four error measures that are defined by the root mean squared error of absolute or relative differences of prices or implied volatilities. Among these measures we have identified two groups: Calibrations with respect to relative prices, absolute implied volatilities or relative implied volatilities lead to similar prices of exotic options. Calibrations with respect to absolute prices imply exotics prices that are quite different from the prices of the first group. The price differences increase for longer times to maturity. Moreover, the differences do not decrease in the Bates model although it is an extension of the Heston model with similar qualitative features and a better fit to plain vanilla data. The price differences of exotic options differ also across option types and are bigger for barrier options than for cliquets.

Moreover, we have looked at the model risk for these two option pricing models. Model risk and calibration risk are not independent because model risk is lowest for calibrations with respect to absolute prices and highest for calibrations with respect to relative implied volatilities. As this holds for all considered options model risk seems to be ordered with respect to the error measure used in the calibration.

As model risk is bigger than calibration risk calibrations should be carried out with respect to absolute prices if the choice of an appropriate model is unclear. But if a model has already been chosen we suggest to measure the error between the model and the market in terms of (relative) implied volatilities because this error measure reflects best the characteristics of the model that are essential for exotic options. Moreover, we have demonstrated that this choice leads to good calibrations (e.g. relatively good fits and stable parameters). We have also shown that the resulting prices of exotic options often lie in the middle of the prices from the other calibrations and have the smallest variance. Our results underline the importance of the implied volatility surface and suggest that one should measure the error in the calibration in terms of implied volatilities.

Chapter 5

Empirical Pricing Kernels and Investor Preferences

So far, we have discussed the market for European options, modelling approaches for option prices and their calibration to market data. In this section, we apply these methods and analyze empirical market utility functions implied by a representative investor. Considering three market regimes in 2000, 2002 and 2004 – a bullish, a bearish and a sidewards market – we estimate the pricing kernels and the corresponding market utility functions from data on the DAX and on options on the DAX. To this end, we use a consistent parametric framework of stochastic volatility but we check the robustness of our results in other – among them nonparametric – models. The empirical market utility functions all show a region of risk proclivity that we reconstruct by considering individual investors whose utility functions have switching point between bullish and bearish attitudes. The inverse problem of finding the distribution of individual switching points is carried out in the space of stock returns by discretization to a quadratic optimization problem.

5.1 Introduction

Numerous attempts have been undertaken to describe basic principles on which the behaviour of individuals are based. Expected utility theory was originally proposed by J. Bernoulli in 1738. In his work J. Bernoulli used such terms as risk aversion and risk premium and proposed a concave (logarithmic) utility function, see Bernoulli (1956). The utilitarianism theory that emerged in the 18th century considered utility maximization as a principle for the organisation of society. Later expected utility was applied to game theory and its properties were formalised by von Neumann and Morgenstern (1944). A

utility function relates some observable variable, in most cases consumption, and an unobservable utility level that this consumption delivers. It was suggested that individuals' preferences are based on this unobservable utility: such bundles of goods are preferred that are associated with higher utility levels. It was claimed that three types of utility functions – concave, convex and linear – correspond to three types of individuals – risk averse, risk neutral and risk seeking. A typical economic agent was considered to be risk averse and this was quantified by coefficients of relative or absolute risk aversion. Another important step in the development of utility theory was the prospect theory of Kahneman and Tversky (1979). By behavioural experiments they found that people act risk averse above a certain reference point and risk seeking below it. This implies that a concave form of the utility function above the reference point and a convex form below it.

Besides these individual utility functions, market utility functions have been analyzed in empirical studies recently by Jackwerth (2000), Rosenberg and Engle (2002) and others. Across different markets, the authors observed a common pattern of market utility functions: There is a reference point near the initial wealth and in a region around this reference point the market utility functions are convex. But for big losses or gains they show a concave form – risk aversion. Such utility functions disagree with the classical utility functions of von Neumann and Morgenstern (1944) and also with the findings of Kahneman and Tversky (1979). They are however in concordance with the utility function form proposed by Friedman and Savage (1948).

In this section, we analyze how these market utility functions can be explained by aggregating individual investors' attitudes. To this end, we first determine empirical pricing kernels from DAX data. Our estimation procedure is based on historical and risk neutral densities and these distributions are derived in stochastic volatility models that are widely used in industry. From these pricing kernels we construct the corresponding market utility functions. Then we describe our method to aggregate individual utility functions to a market utility function. This leads to an inverse problem for a density function that describes how many investors have a utility function of a special type. We solve this problem by a discrete approximation. In this way, we derive utility functions and their distribution among investors that allow to recover the market utility function. Hence, we explain how individual utility functions can be used to form the behaviour of the whole market.

In section 5.2, we describe the theoretical connection between utility functions and pricing kernels. In section 5.3, we present a consistent stochastic volatility framework for the estimation of both the historical and the risk neutral density. Moreover, we discuss the empirical pricing kernel implied

by the DAX in 2000, 2002 and 2004. In section 5.4, we explain the utility aggregation method that relates the market utility function and the utility functions of individual investors. This aggregation mechanism leads to an inverse problem that is analyzed and solved in this section. In section 5.5, we conclude and discuss related approaches.

5.2 Pricing kernels and utility functions

In this section, we derive the fundamental relationship between utility functions and pricing kernels. It describes how a representative utility function can be derived from historical and risk-neutral distributions of assets. In the following sections, we estimate the empirical pricing kernel and observe in this way the market utility function.

First, we derive the price of a security in an equilibrium model: We consider an investor with a utility function U who has as initial endowment one share of stock. She can invest into the stock and a bond up to a final time when she can consume. Her problem is to choose a strategy that maximizes the expected utility of her initial and terminal wealth. In a diffusion framework, this leads to a well known optimization problem introduced by Merton (1971). The solution for the investor is to hold the stock all the time so that her terminal wealth is given by the stock.

From this result, we can derive the asset pricing equation

$$P_0 = \mathrm{E}^P \left[\psi(S_T) M_T\right] \tag{5.1}$$

for a security on the stock (S_t) with payoff function ψ at maturity T. Here, P_0 denotes the price of the security at time 0 and E is the expectation with respect to the real/historical measure P. The stochastic discount factor M_T is given by

$$M_T = \beta U'(S_T)/U'(S_0)$$

where β is a fixed discount factor. This stochastic discount factor is actually the projection of the general stochastic discount factor on the traded asset (S_t). The stochastic discount factor can depend on more variables in general. But as discussed in Cochrane (2001) this projection has the same interpretation for pricing as the general stochastic discount factor.

Besides this equilibrium based approach, Black and Scholes (1973) derived the price of a security relative to the underlying by constructing a

perfect hedge. The resulting continuous delta hedging strategy is equivalent to pricing under a risk neutral measure Q under which the discounted price process of the underlying becomes a martingale. Hence, the price of a security is given by an expected value with respect to a risk neutral measure Q:

$$P_0 = \mathrm{E}^Q \left[\exp(-rT)\psi(S_T)\right]$$

If p denotes the historical density of S_T (i.e. $P(S_T \leq s) = \int_{-\infty}^{s} p(x)\ dx$) and q the risk neutral density of S_T (i.e. $Q(S_T \leq s) = \int_{-\infty}^{s} q(x)\ dx$) then we get

$$\begin{aligned} P_0 &= \exp(-rT) \int \psi(x) q(x) dx \\ &= \exp(-rT) \int \psi(x) \frac{q(x)}{p(x)} p(x) dx \\ &= \mathrm{E}^P \left[\exp(-rT)\psi(S_T)\frac{q(S_T)}{p(S_T)}\right] \end{aligned} \tag{5.2}$$

As equations (5.1) and (5.2) hold for all payoff profiles we can conclude

$$\beta \frac{U'(s)}{U'(S_0)} = \exp(-rT)\frac{q(s)}{p(s)}.$$

Defining the pricing kernel by $\mathcal{K} = q/p$ we conclude that the form of the market utility function can be derived from the empirical pricing kernel by integration:

$$\begin{aligned} U(s) &= U(S_0) + \int_{S_0}^{s} U'(S_0) \frac{\exp(-rT)}{\beta} \frac{q(x)}{p(x)} dx \\ &= U(S_0) + \int_{S_0}^{s} U'(S_0) \frac{\exp(-rT)}{\beta} \mathcal{K}(x) dx \end{aligned}$$

because S_0 is known.

As an example, we consider the model of Black and Scholes (1973) where the stock follows a geometric Brownian motion

$$dS_t/S_t = \mu dt + \sigma dW_t \tag{5.3}$$

Here the historical density p of S_t is log-normal, i.e.

$$p(x) = \frac{1}{x}\frac{1}{\sqrt{2\pi\tilde{\sigma}^2}} \exp\left\{-\frac{1}{2}\left(\frac{\log x - \tilde{\mu}}{\tilde{\sigma}}\right)^2\right\},\ x > 0$$

where $\tilde{\mu} = (\mu - \sigma^2/2)t + \log S_0$ and $\tilde{\sigma} = \sigma\sqrt{t}$. Under the risk neutral measure Q the drift μ is replaced by the riskless interest rate r, see e.g. Harrison and Pliska (1981). Thus, also the risk neutral density q is log-normal. In this way, we can derive the pricing kernel

$$\mathcal{K}(x) = \left(\frac{x}{S_0}\right)^{-\frac{\mu-r}{\sigma^2}} \exp\{(\mu - r)(\mu + r - \sigma^2)T/(2\sigma^2)\}.$$

This pricing kernel has the form of a derivative of a power utility

$$\mathcal{K}(x) = \lambda \left(\frac{x}{S_0}\right)^{-\gamma}$$

where the constants are given by $\lambda = e^{\frac{(\mu-r)(\mu+r-\sigma^2)T}{2\sigma^2}}$ and $\gamma = \frac{\mu-r}{\sigma^2}$. This gives a utility function corresponding to the underlying (5.3)

$$U(S_T) = (1 - \frac{\mu - r}{\sigma^2})^{-1} \; S_T^{(1-\frac{\mu-r}{\sigma^2})}$$

where we ignored additive and multiplicative constants. In this power utility function the risk aversion is not given by the market price of risk $(\mu - r)/\sigma$. Instead investors take the volatility more into account. The expected return $\mu - r$ that is adjusted by the riskfree return is related to the variance. This results in a higher relative risk aversion than the market price of risk.

A utility function corresponding to the Black-Scholes model is shown in the upper panel of figure 5.1 as a function of returns. In order to make different market situations comparable we consider utility functions as functions of (half year) returns $R = S_{0.5}/S_0$. We chose the time horizon of half a year ahead for our analysis. Shorter time horizons are interesting economically and moreover the historical density converges to the Dirac measure so that results become trivial (in the end). Longer time horizons are economically more interesting but it is hardly possible to estimate the historical density for a long time ahead. It does not seem neither realistic to assume that investors have clear ideas where the DAX will be in e.g. 10 years. For these reasons we use half a year as future horizon. Utility functions $\tilde{U}$ of returns are defined by:

$$\tilde{U}(R) := U(RS_0), \;\; R > 0$$

where S_0 denotes the value of the DAX on the day of estimation. Because of $U' = c\mathcal{K}$ for a constant c we have $\tilde{U}'(R) = c\mathcal{K}(RS_0)S_0$ and we see that also utility functions of returns are given by integrals of the pricing kernel. The

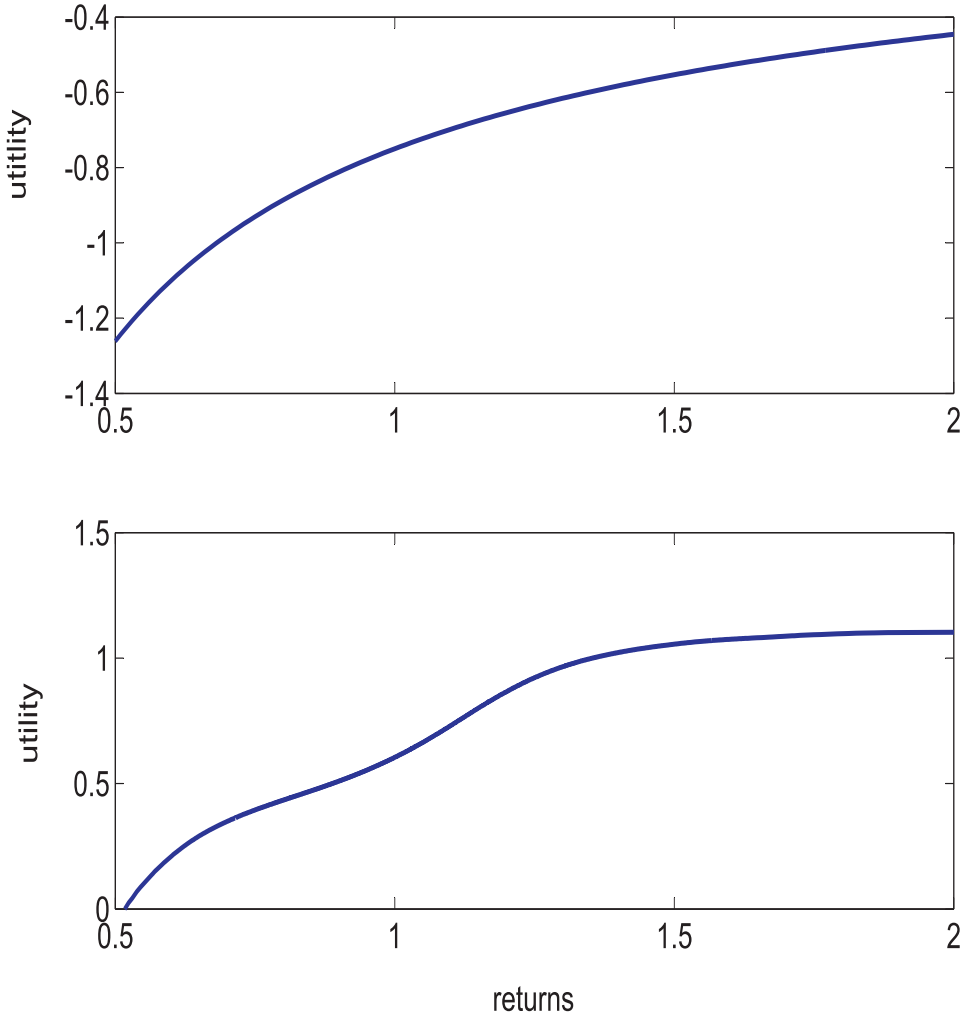

Figure 5.1: up: Utility function in the Black Scholes model for $T = 0.5$ years ahead and drift $\mu = 0.1$, volatility $\sigma = 0.2$ and interest rate $r = 0.03$. down: Market utility function on 06/30/2000 for $T = 0.5$ years ahead.

change to returns allows us to compare different market regimes independently of the initial wealth. In the following we denote the utility functions of returns by the original notation U. Hence, we suppress in the notation the dependence of the utility function U on the day of estimation t.

The utility function corresponding to the model of Black and Scholes (1973) is a power utility, monotonically increasing and concave. But such classical utility functions are not observed on the market. Parametric and nonparametric models that replicate the option prices all lead to utility functions with a hump around the initial wealth level. This is described in detail later but is shown already in figure 5.1. The upper panel presents the utility function corresponding to Black-Scholes model with a volatility of 20% and an expected return of 10%. The function is concave and implies a constant relative risk aversion. The utility function estimated on the bullish market in summer 2000 is presented in the lower panel. Here, the hump around the money is clearly visible. The function is no more concave but has a region where investors are risk seeking. This risk proclivity around the money is reflected in a negative relative risk aversion.

5.3 Estimation

In this section, we start by reviewing some recent approaches for estimating the pricing kernel. Then we describe our method that is based on estimates of the risk neutral and the historical density. The risk neutral density is derived from option prices that are given by an implied volatility surface and the historical density is estimated from the independent data set of historical returns. Finally, we present the empirical pricing kernels and the inferred utility and relative risk aversion functions.

5.3.1 Estimation approaches for the pricing kernel

There exist several ways and methods to estimate the pricing kernel. Some of these methods assume parametric models while others use nonparametric techniques. Moreover, some methods estimate first the risk neutral and subjective density to infer the pricing kernel. Other approaches estimate directly the pricing kernel.

Ait-Sahalia and Lo (1998) derive a nonparametric estimator of the risk neutral density based on option prices. In Ait-Sahalia and Lo (2000), they consider the empirical pricing kernel and the corresponding risk aversion

using this estimator. Moreover, they derive asymptotic properties of the estimator that allow e.g. the construction of confidence bands. The estimation procedure consists of two steps: First, the option price function is determined by nonparametric kernel regression and then the risk neutral density is computed by the formula of Breeden and Litzenberger (1978). Advantages of this approach are the known asymptotic properties of the estimator and the few assumptions necessary.

Jackwerth (2000) analyses risk aversion by computing the risk neutral density from option prices and the subjective density from historical data of the underlying. For the risk neutral distribution, he applies a variation of the estimation procedure described in Jackwerth and Rubinstein (1996): A smooth volatility function derived from observed option prices gives the risk neutral density by differentiating it twice. The subjective density is approximated by a kernel density computed from historical data. In this method bandwidths have to be chosen as in the method of Ait-Sahalia and Lo (1998).

Rosenberg and Engle (2002) use a different approach and estimate the subjective density and directly (the projection of) the pricing kernel. This gives the same information as the estimation of the two densities because the risk neutral density is the product of the pricing kernel and the subjective density. For the pricing kernel, they consider two parametric specifications as power functions and as exponentials of polynomials. The evolution of the underlying is modelled by GARCH processes. As the parametric pricing kernels lead to different results according to the parametric form used this parametric approach appears a bit problematic.

Chernov (2003) also estimates the pricing kernel without computing the risk neutral and subjective density explicitly. Instead of assuming directly a parametric form of the kernel he starts with a (multi dimensional) modified model of Heston (1993) and derives an analytic expression for the pricing kernel by the Girsanov theorem, see Chernov (2000) for details. The kernel is estimated by a simulated method of moments technique from equity, fixed income and commodities data and by reprojection. An advantage of this approach is that the pricing kernel is estimated without assuming an equity index to approximate the whole market portfolio. But the estimation procedure is rather complex and model dependent.

In a recent paper, Barone-Adesi et al. (2004) price options in a GARCH framework allowing the volatility to differ between historical and risk neutral distribution. This approach leads to acceptable calibration errors between the observed option prices and the model prices. They estimate the historical density as a GARCH process and consider the pricing kernel only on one day. This kernel is decreasing which coincides with standard economic the-

ory. But the general approach of changing explicitly the volatility between the historical and risk neutral distribution is not supported by the standard economic theory.

We estimate the pricing kernel by estimating the risk neutral and the subjective density and then deriving the pricing kernel. This approach does not impose a strict structure on the kernel. Moreover, we use accepted parametric models because nonparametric techniques for the estimation of second derivatives depend a lot on the bandwidth selection although they yield the same pricing kernel behaviour over a wide range of bandwidths. For the risk neutral density we use a stochastic volatility model that is popular both in academia and in industry. The historical density is more difficult to estimate because the drift is not fixed. Hence, the estimation depends more on the model and the length of the historical time series. In order to get robust results we consider different (discrete) models and different lengths. In particular, we use a GARCH model that is the discrete version of the continuous model for the risk neutral density. In the following, we describe these models, their estimation and the empirical results.

5.3.2 Estimation of the risk neutral density

Stochastic volatility models are popular in industry because they replicate the observed smile in the implied volatility surfaces (IVS) rather well and moreover imply rather realistic dynamics of the surfaces. Nonparametric approaches like the local volatility model of Dupire (1994) allow a perfect fit to observed price surfaces but their dynamics are in general contrary to the market. As Bergomi (2005) points out the dynamics are more important for modern products than a perfect fit. Hence, stochastic volatility models are popular.

We consider the model of Heston (1993) for the risk neutral density because it can be interpreted as the limit of GARCH models. The Heston model has been refined further in order to improve the fit, e.g. by jumps in the stock price or by a time varying mean variance level. We use the original Heston model in order to maintain a direct connection to GARCH processes. Although it is possible to estimate the historical density also with the Heston model e.g. by Kalman filter methods we prefer more direct approaches in order to reduce the dependence of the results on the model and the estimation technique.

The stochastic volatility model of Heston (1993) is given by the two

stochastic differential equations:

$$\frac{dS_t}{S_t} = rdt + \sqrt{V_t}dW_t^1$$

where the variance process is modelled by a square-root process:

$$dV_t = \xi(\eta - V_t)dt + \theta\sqrt{V_t}dW_t^2$$

and W^1 and W^2 are Wiener processes with correlation ρ and r is the risk free interest rate. The first equation models the stock returns by normal innovations with stochastic variance. The second equation models the stochastic variance process as a square-root diffusion.

The parameters of the model all have economic interpretations: η is called the long variance because the process always returns to this level. If the variance V_t is e.g. below the long variance then $\eta - V_t$ is positive and the drift drives the variance in the direction of the long variance. ξ controls the speed at which the variance is driven to the long variance. In calibrations, this parameter changes a lot and makes also the other parameters instable. To avoid this problem, the reversion speed is kept fixed in general. We follow this approach and choose $\xi = 2$ as Bergomi (2005) does. The volatility of variance θ controls mainly the kurtosis of the distribution of the variance. Moreover, there are the initial variance V_0 of the variance process and the correlation ρ between the Brownian motions. This correlation models the leverage effect: When the stock goes down then the variance goes up and vice versa. The parameters also control different aspects of the implied volatility surface. The short (long) variance determines the level of implied volatility for short (long) maturities. The correlation creates the skew effect and the volatility of variance controls the smile.

The variance process remains positive if the volatility of variance θ is small enough with respect to the product of the mean reversion speed ξ and the long variance level η (i.e. $2\xi\eta > \theta^2$). As this constraint leads often to significantly worse fits to implied volatility surfaces it is in general not taken into account and we follow this approach.

The popularity of this model can probably be attributed to the semiclosed form of the prices of plain vanilla options. Carr and Madan (1999) showed that the price $C(K,T)$ of a European call option with strike K and maturity T is given by

$$C(K,T) = \frac{\exp\{-\alpha\ln(K)\}}{\pi}\int_0^{+\infty}\exp\{-\mathbf{i}v\ln(K)\}\psi_T(v)dv$$

for a (suitable) damping factor $\alpha > 0$. The function ψ_T is given by

$$\psi_T(v) = \frac{\exp(-rT)\phi_T\{v-(\alpha+1)\mathbf{i}\}}{\alpha^2+\alpha-v^2+\mathbf{i}(2\alpha+1)v}$$

where ϕ_T is the characteristic function of $\log(S_T)$. This characteristic function is given by

$$\begin{aligned}\phi_T(z) &= \exp\{\frac{-(z^2+\mathbf{i}z)V_0}{\gamma(z)\coth\frac{\gamma(z)T}{2}+\xi-\mathbf{i}\rho\theta z}\} \\ &\quad\times\frac{\exp\{\frac{\xi\eta T(\xi-\mathbf{i}\rho\theta z)}{\theta^2}+\mathbf{i}zTr+\mathbf{i}z\log(S_0)\}}{(\cosh\frac{\gamma(z)T}{2}+\frac{\xi-\mathbf{i}\rho\theta z}{\gamma(z)}\sinh\frac{\gamma(z)T}{2})^{\frac{2\xi\eta}{\theta^2}}}\end{aligned} \tag{5.4}$$

where $\gamma(z) \stackrel{\text{def}}{=} \sqrt{\theta^2(z^2+\mathbf{i}z)+(\xi-\mathbf{i}\rho\theta z)^2}$, see e.g. Cizek et al. (2005).

For the calibration we minimize the absolute error of implied volatilities based on the root mean square error:

$$\text{ASE}_t \stackrel{\text{def}}{=} \sqrt{\sum_{i=1}^{n} n^{-1}\{IV_i^{mod}(t)-IV_i^{mar}(t)\}^2}$$

where *mod* refers to a model quantity, *mar* to a quantity observed on the market and $IV(t)$ to an implied volatility on day t. The index i runs over all n observations of the surface on day t.

It is essential for the error functional ASE_t which observed prices are used for the calibration. As we investigate the pricing kernel for half a year to maturity we use only the prices of options that expire in less than 1.5 years. In order to exclude liquidity problems occurring at expiry we consider for the calibration only options with more than 1 month time to maturity. In the moneyness direction we restrict ourselves to strikes 50% above or below the spot for liquidity reasons.

The risk neutral density is derived by estimation of the model parameters by a least squares approach. This amounts to the minimization of the error functional ASE_t. Cont and Tankov (2004b) provided evidence that such error functionals may have local minima. In order to circumvent this problem we apply a stochastic optimization routine that does not get trapped in a local minimum. To this end, we use the method of differential evolution developed by Storn and Price (1997).

Having estimated the model parameters we know the distribution of $X_T = \log S_T$ in form of the characteristic function ϕ_T, see (5.4). Then the

corresponding density f of X_T can be recovered by Fourier inversion:

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{\mathbf{i}tx} \phi_T(t) dt,$$

see e.g. Billingsley (1995). This integral can be computed numerically.

Finally, the risk neutral density q of $S_T = \exp(X_T)$ is given as a transformed density:

$$q(x) = \frac{1}{x} f\{\log(x)\}.$$

This density q is risk neutral because it is derived from option prices and options are priced under the risk neutral measure. This measure is applied because banks replicate the payoff of options so that no arbitrage conditions determine the option price, see e.g. Rubinstein (1994). An estimated risk neutral density is presented in figure 5.2. It is estimated from the implied volatility shown in figure 5.3 for the day 24/03/2000. The distribution is right skewed and its mean is fixed by the martingale property. This implies that the density is low for high profits and high for high losses. Moreover, the distribution is not symmetrical around the neutral point where there are neither profits nor losses. For this and all the following estimations we approximate the risk free interest rates by the EURIBOR. On each trading day we use the yields corresponding to the maturities of the implied volatility surface. As the DAX is a performance index it is adjusted to dividend payments. Thus, we do not have to consider dividend payments explicitly.

5.3.3 Estimation of the historical density

While the risk neutral density is derived from option prices observed on the day of estimation we derive the subjective density from the historical time series of the index. Hence, the two data sets are independent in the sense that the option prices reflect the future movements and the historical time series the past.

The estimation of the historical density seems more difficult than the estimation of the risk neutral density because the drift is not fixed and it depends in general on the length of the time series. Because of these difficulties we use different models and time horizons for the historical density: First, we estimate a GARCH in mean model for the returns. Returns are generally assumed to be stationary and we confirmed this at least in the time intervals we consider. The mean component in the GARCH model is important to reflect different market regimes. We estimate the GARCH model from the time series of the returns of the last two year because GARCH models require

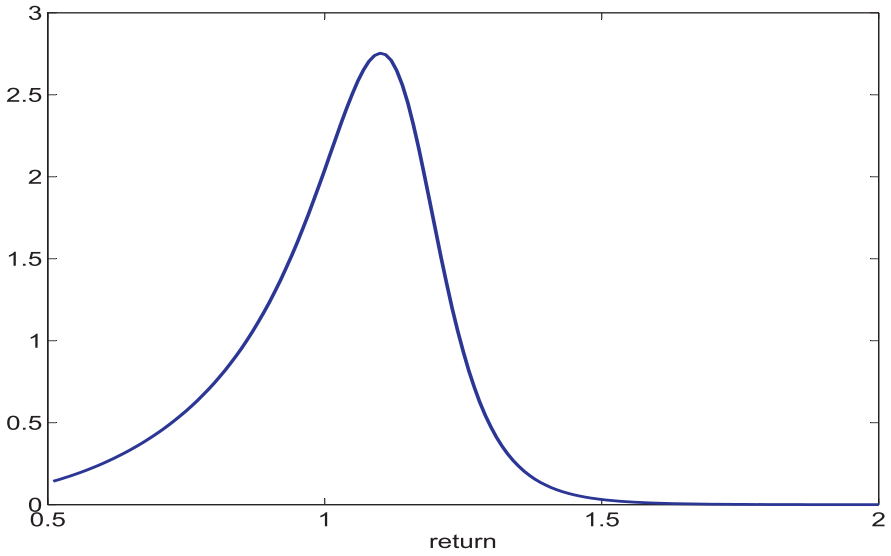

Figure 5.2: Risk neutral density on 24/03/2000 half a year ahead.

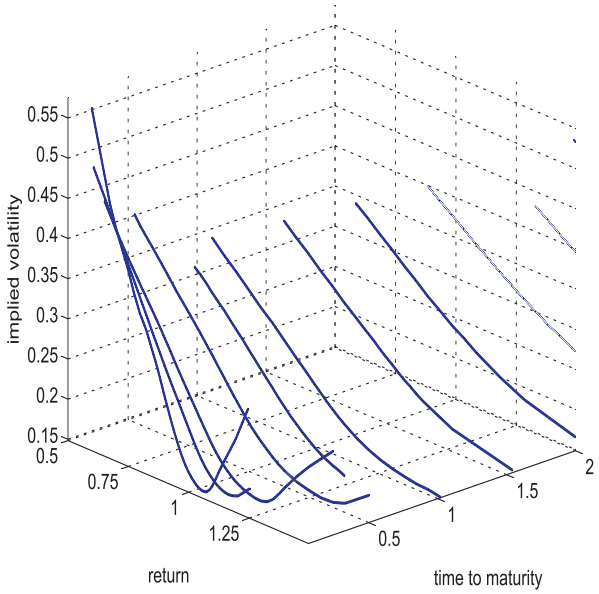

Figure 5.3: Implied volatility surface on 24/03/00.

model	time period
GARCH in mean	2.0y
discrete Heston	2.0y
observed returns	1.0y

Table 5.1: Models and the time periods used for their estimation.

quite long time series for the estimation in order to make the standard error reasonably small. We do not choose longer time period for the estimation because we want to consider special market regimes. Besides this popular model choice we apply a GARCH model that converges in the limit to the Heston model that we used for the risk neutral density. As this model is also hard to estimate we use again the returns of the last 2 years for this model. Moreover, we consider directly the observed returns of the last year. The models and their time period for the estimation are presented in table 5.1. All these models give by simulation and smoothing the historical density for half a year ahead.

The GARCH estimations are based on the daily log-returns

$$R_i = \log(S_{t_i}) - \log(S_{t_{i-1}})$$

where (S_t) denotes the price process of the underlying and $t_i,\ i = 1, 2, \ldots$ denote the settlement times of the trading days. Returns of financial assets have been analyzed in numerous studies, see e.g. Cont (2001). A model that has often been successfully applied to financial returns and their stylized facts is the GARCH(1,1) model. This model with a mean is given by

$$\begin{aligned} R_i &= \mu + \sigma_i Z_i \\ \sigma_i^2 &= \omega + \alpha R_{i-1}^2 + \beta \sigma_{i-1}^2 \end{aligned}$$

where (Z_i) are independent identically distributed innovations with a standard normal distribution, see e.g. Franke et al. (2004). On day t_j the model parameters μ, ω, α and β are estimated by quasi maximum likelihood from the observations of the last two years, i.e. $R_{j-504}, \ldots, R_j$ assuming 252 trading days per year.

After the model parameters have been estimated on day t_j from historical data the process of logarithmic returns (R_i) is simulated half a year ahead, i.e. until time $t_j + 0.5$. In such a simulation μ, ω, α and β are given and the time series (σ_i) and (R_i) are unknown. The values of the DAX corresponding to the simulated returns are then given by inverting the definition of the log

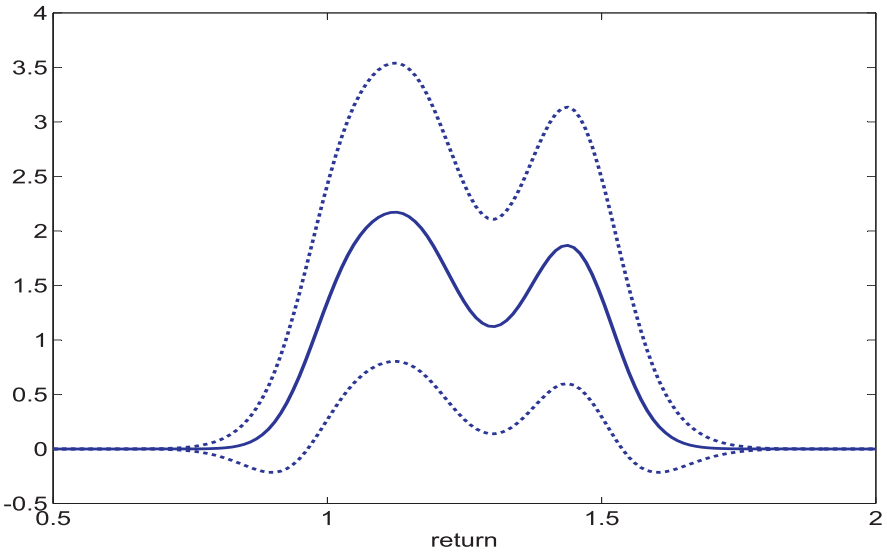

Figure 5.4: Historical density on 24/03/2000 half a year ahead with 95% confidence band.

returns:

$$S_{t_i} = S_{t_{i-1}} \exp(R_i)$$

where we start with the observed DAX value on day t_j. Repeating the simulation N times we obtain N samples of the distribution of $S_{t_j+0.5}$. We use $N = 2000$ simulations because tests have shown that the results become robust around this number of simulations.

From these samples we estimate the probability density function of $S_{t_j+0.5}$ (given $(S_{t_{j-126}}, \ldots, S_{t_j})$) by kernel density estimation. We apply the Gaussian kernel and choose the bandwidth by Silverman's rule of thumb, see e.g. Silverman (1986). This rule provides a trade-off between oversmoothing – resulting in a high bias – and undersmoothing – leading to big variations of the density. We have moreover checked the robustness of the estimate relative to this bandwidth choice. The estimation results of a historical density are presented in figure 5.4 for the day 24/03/2000. This density that represents a bullish market is has most of its weight in the profit region and its tail for the losses is relatively light.

As we use the Heston model for the estimation of the risk neutral density we consider in addition to the described GARCH model a GARCH model that is a discrete version of the Heston model. Heston and Nandi (2000)

show that the discrete version of the square-root process is given by

$$V_i = \omega + \beta V_{i-1} + \alpha(Z_{i-1} - \gamma\sqrt{V_{i-1}})$$

and the returns are modelled by

$$R_i = \mu - \frac{1}{2}V_i + \sqrt{V_i}Z_i$$

where (Z_i) are independent identically distributed innovations with a standard normal distribution. Having estimated this model by maximum likelihood on day t_j we simulate it half a year ahead and then smooth the samples of $S_{t_j+0.5}$ in the same way as in the other GARCH model.

In addition to these parametric models, we consider directly the observed returns over half a year

$$\tilde{R}_i = S_{t_i}/S_{t_{i-126}}.$$

In this way, we interpret these half year returns as samples from the distribution of the returns for half a year ahead. Smoothing these historical samples of returns gives an estimate of the density of returns and in this way also an estimate of the historical density of $S_{t_j+0.5}$.

5.3.4 Empirical pricing kernels

In contrast to many other studies that concentrate on the S&P500 index we analyze the German economy by focusing on the DAX, the German stock index. This broad index serves as an approximation to the German economy. We use two data sets: A daily time series of the DAX for the estimation of the subjective density and prices of European options on the DAX for the estimation of the risk neutral density.

In figure 5.5, we present the DAX in the years 1998 to 2004. This figure shows that the index reached its peak in 2000 when all the internet firms were making huge profits. But in the same year this bubble burst and the index fell afterwards for a long time. The historical density is estimated from the returns of this time series. We analyze the market utility functions in March 2000, July 2002 and June 2004 in order to consider different market regimes. We interpret 2000 as a bullish, 2002 as a bearish and 2004 as a sidewards market. These interpretations are based on table 5.2 that describes the changes of the DAX over the preceding 1 or 2 years. (In June 2004 the market went up by 11% in the last 10 months.)

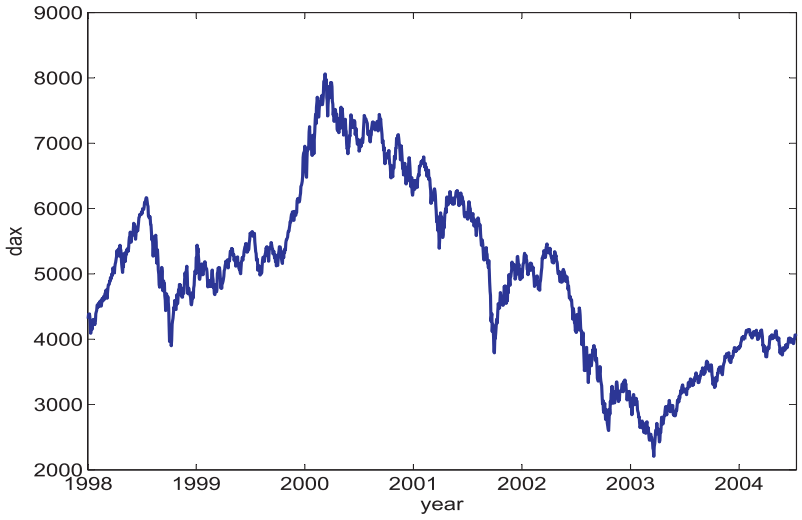

Figure 5.5: DAX, 1998 - 2004.

	1.0y	2.0y
03/2000	1.63	1.57
07/2002	0.66	0.54
06/2004	1.11	0.98

Table 5.2: Market regimes in 2000, 2002 and 2004 described by the return $S_0/S_{0-\Delta}$ for periods $\Delta = 1.0y, 2.0y$.

A utility function derived from the market data is a market utility function. It is estimated as an aggregate for all investors as if the representative investor existed. A representative investor is however just a convenient construction because the existence of the market itself implies that the asset is bought and sold, i.e. at least two counterparties are required for each transaction.

In section 5.2 we identified the market utility function (up to linear transformations) as

$$U(R) = \int_1^R \mathcal{K}(x)dx$$

where $\mathcal{K}$ is the pricing kernel for returns. It is defined by

$$\mathcal{K}(x) = q(x)/p(x)$$

in terms of the historical and risk neutral densities p and q of returns. Any utility function (both cardinal and ordinal) can be defined up to a linear transformation, therefore we have identified the utility functions sufficiently. In section 5.3.3 we proposed different models for estimating the historical density. In figure 5.6 we show the pricing kernels resulting from the different estimation approaches for the historical density. The figure shows that all three kernels are quite similar: They have the same form, the same characteristic features like e.g. the hump and differ in absolute terms only a little. This demonstrates the economic equivalence of the three estimation methods on this day and this equivalence holds also for the other days. In the following we work with historical densities that are estimated by the observed returns.

Besides the pricing kernel and the utility function we consider also the risk attitudes in the markets. Such risk attitudes are often described in terms of relative risk aversion that is defined by

$$RRA(R) = -R\frac{U''(R)}{U'(R)}.$$

Because of $U' = c\mathcal{K} = cq/p$ for a constant c the relative risk aversion is also given by

$$RRA(R) = -R\frac{q'(R)p(R) - q(R)p'(R)}{p^2(R)} \Big/ \frac{q(R)}{p(R)} = R\left(\frac{p'(R)}{p(R)} - \frac{q'(R)}{q(R)}\right).$$

Hence, we can estimate the relative risk aversion from the estimated historical and risk neutral densities.

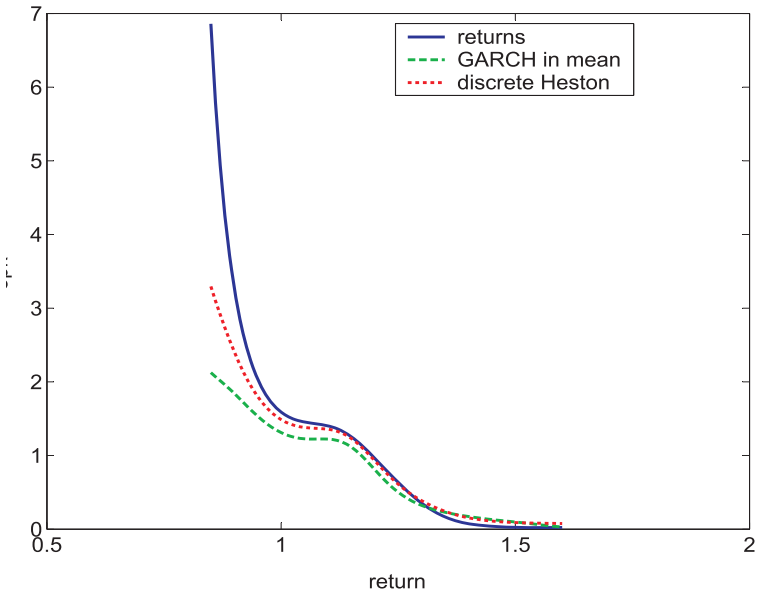

Figure 5.6: Empirical pricing kernel on 24/03/2000.

In figure 5.7 we present the empirical pricing kernels in March 2000, July 2002 and June 2004. The dates represent a bullish, a bearish and a sidewards markets, see table 5.2. All pricing kernels have a proclaimed hump located at small profits. Hence, the market utility functions do not correspond to standard specification of utility functions. We present the pricing kernels only in regions around the initial DAX (corresponding to a return of 1) value because the kernels explode outside these regions. This explosive behaviour reflects the typical pricing kernel form for losses. The explosion of the kernel for large profits is due to numerical problems in the estimation of the very low densities in this region. But we can see that in the sidewards market the kernel is concentrated on a small region while the bullish and bearish markets have wider pricing kernels. The hump of the sidewards market is also narrower than in the other two regimes. The bullish and bearish regimes have kernels of similar width but the bearish kernel is shifted to the loss region and the bullish kernel is located mainly in the profit area. Moreover, the figures show that the kernel is steeper in the sidewards markets than in the other markets. But this steepness cannot be interpreted clearly because pricing kernels are only defined up to a multiplicative constant.

The pricing kernels are the link between the relative risk aversion and the utility functions that are presented in figure 5.8. These utility functions are only defined up to linear transformations, see section 5.2. All the utility

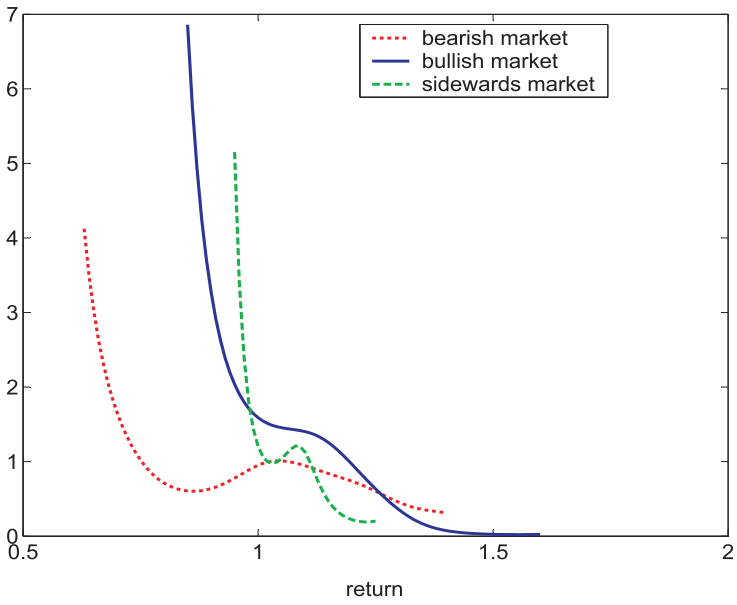

Figure 5.7: Empirical pricing kernel on 24/03/2000, 30/07/2002 and 30/06/2004.

functions are increasing but only the utility function of the bullish market is concave. This concavity can be seen from the monotonicity of the kernel, see figure 5.7. Actually, this non convexity can be attributed to the quite special form of the historical density which has two modes on this date, see figure 5.4. Hence, we presume that also this utility function has in general a region of convexity. The other two utility functions are convex in a region of small profits where the bullish utility is almost convex. The derivatives of the utility functions cannot be compared directly because utility functions are identified only up to multiplicative constants. But we can compare the ratio of the derivatives in the loss and profit regions for the three dates because the constants cancel in these ratios. We see that the derivatives in the loss region are highest in the bullish and lowest in the bearish market and vice versa in the profit region. Economically these observations can be interpreted in such a way that in the bullish market a loss (of 1 unit) reduces the utility stronger than in the bearish market. On the other hand, a gain (of 1 unit) increases the utility less than in the bearish market. The sidewards market shows a behaviour between these extreme markets. Hence, investors fear in a good market situation losses more than in a bad situation and they appreciate profits in a good situation less than in a bad situation.

Finally, we consider the relative risk aversions in the three market regimes.

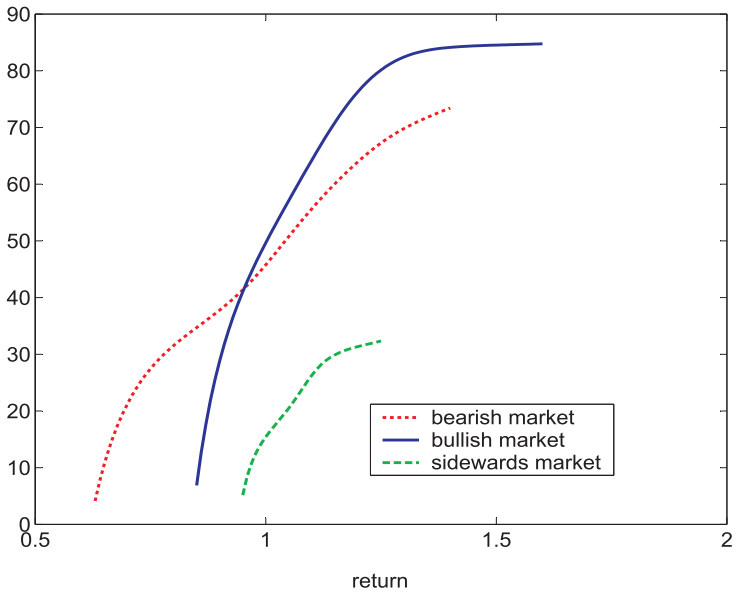

Figure 5.8: Market utility functions on 24/03/2000, 30/07/2002 and 30/06/2004.

These risk aversions are presented in figure 5.9, they do not depend on any constants but are completed identified. We see that the risk aversion is smallest in all markets for a small profit that roughly corresponds to the initial value plus a riskless interest on it. In the sidewards regime the market is risk seeking in a small region around this minimal risk aversion. But then the risk aversion increases quite fast. Hence, the representative agent in this market is willing to take small risks but is sensitive to large losses or profits. In the bullish and bearish regimes the representative agent is less sensitive to large losses or profits than in the sidewards market. In the bearish situation the representative agent is willing to take more risks than in the bullish regime. In the bearish regime the investors are risk seeking in a wider region than in the sidewards regime. In this sense they are more risk seeking in the bearish market. In the bullish market – on the other hand – the investors are never risk seeking so that they are less risk seeking than in the sidewards market.

The estimated utility functions most closely follow the specification proposed by Friedman and Savage (1948). The utility function proposed by Kahneman and Tversky (1979) consists of one concave and one convex segment and is less suitable for describing the observed behaviour, see figure 5.10. Both utility functions were proposed to account for two opposite types

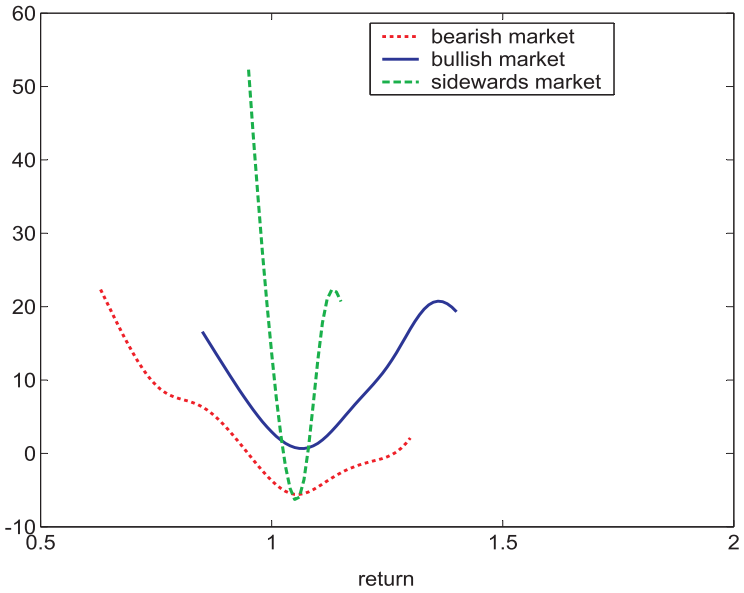

Figure 5.9: Relative risk aversions on 24/03/2000, 30/07/2002 and 30/06/2004.

of behaviour with respect to risk attitudes: buying insurance and gambling. Any utility function that is strictly concave fails to describe both risk attitudes. Most notable examples are the quadratic utility function with the linear pricing kernel as in the CAPM model and the CRRA utility function. These functions are presented in figure 5.10. Comparing this theoretical figure with the empirical results in figure 5.7 we see clearly the shortcoming of the standard specifications of utility functions to capture the characteristic hump of the pricing kernels.

5.4 Individual investors and their utility functions

In this section, we introduce a type of utility function that has two regions of different risk aversion. Then we describe how individual investors can be aggregated to a representative agent that has the market utility function. Finally, we solve the resulting estimation problem by discretization and estimate the distribution of individual investors.

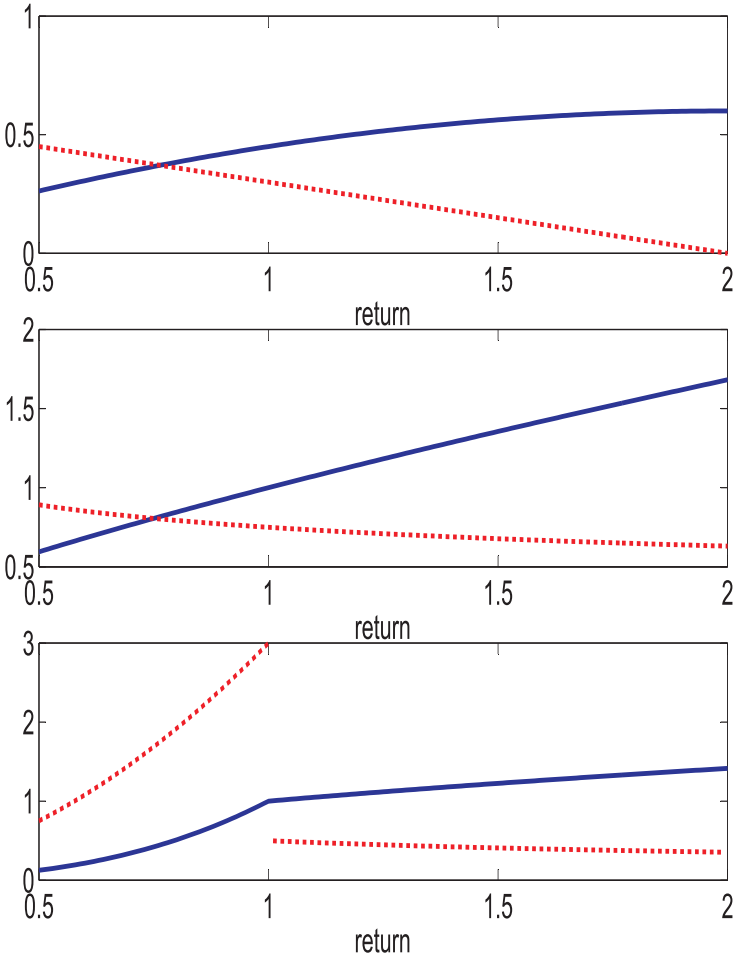

Figure 5.10: Common utility functions (solid) and their pricing kernels (dotted) (upper: quadratic, middle: power, lower panel: Kahneman and Tversky utility function).

5.4.1 Individual Utility Function

We learn from figures 5.10 and 5.7 that the market utility differs significantly from the standard specification of utility functions. Moreover, we can observe from the estimated utility functions 5.8 that the loss part and the profit part of the utility functions can be quite well approximated with shifted CRRA functions, $k = 1, 2$:

$$U^{(k)}(R) = a_k \frac{(R - c_k)^{1-\gamma_k}}{1 - \gamma_k} + b_k,$$

where the shift parameter is c_k. These power utility functions become infinitely negative for $R = c_k$ and can be extended by $U^{(k)}(R) = -\infty$ for $R \leq c_k$, i.e. investors will avoid by all means the situation when $R \leq c_k$. The standard CRRA utility function has $c_k = 0$.

We try to reconstruct the market utility of the representative investor by individual utility functions and hence assume that there are many investors on the market. Investor i will be attributed with a utility function that consists of two CRRA functions:

$$U_i(R) = \begin{cases} \max\left\{U(R, \theta_1, c_1); U(R, \theta_2, c_{2,i})\right\}, & \text{if} \quad R > c_1 \\ -\infty, & \text{if} \quad R \leq c_1 \end{cases}$$

where $U(R, \theta, c) = a\frac{(R-c)^{1-\gamma}}{1-\gamma} + b$, $\theta = (a, b, \gamma)^\top$, $c_{2,i} > c_1$. If $a_1 = a_2 = 1$, $b_1 = b_2 = 0$ and $c_1 = c_2 = 0$, we get the standard CRRA utility function.

The parameters θ_1 and θ_2 and c_1 are the same for all investors who differ only with the shift parameter c_2. θ_1 and c_1 are estimated from a loss part of the utility market function, where all investors probably agree that the market is "bad". θ_2 is estimated from a profit part of the utility function where all investors agree that the state of the world is "good". The distribution of c_2 uniquely defines the distribution of switching points and is computed in section 5.4.3. In this way a bear part $U_{bear}(R) = U(R, \theta_1, c_1)$ and a bull part $U_{bull}(R) = U(R, \theta_1, c_2)$ can be estimated by least squares.

The individual utility function can then be denoted conveniently as:

$$U_i(R) = \begin{cases} \max\left\{U_{bear}(R); U_{bull}(R, c_i)\right\}, & \text{if} \quad R > c_1; \\ -\infty, & \text{if} \quad R \leq c_1. \end{cases} \tag{5.5}$$

Switching between U_{bear} and U_{bull} happens at the *switching point* z, whereas $U_{bear}(z) = U_{bull}(z, c_i)$. The switching point is uniquely determined by $c_i \equiv c_{2,i}$. The notations *bear* and *bull* have been chosen because U_{bear} is activated when returns are low and U_{bull} when returns are high.

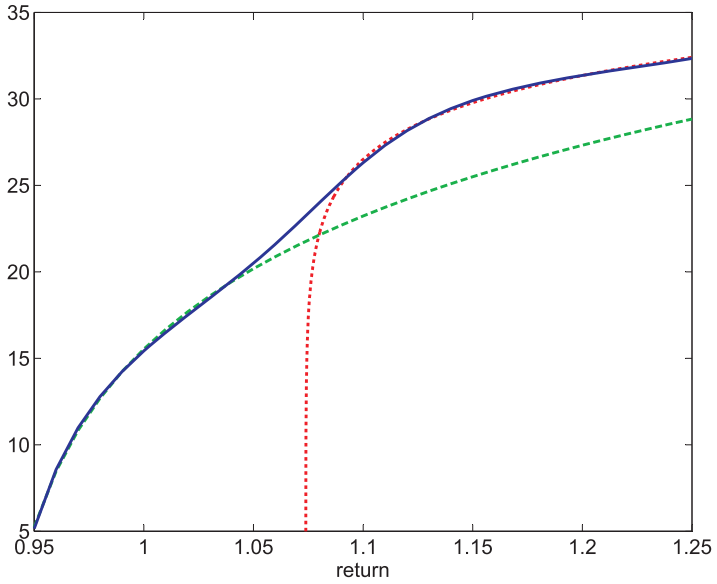

Figure 5.11: Market utility function (solid) with bearish (dashed) and bullish (dotted) part of an individual utility function 5.5 estimated in the sidewards market of 30/06/2004.

Each investor is characterised by a switching point z. The smoothness of the market utility function is the result of the aggregation of different attitudes. U_{bear} characterises more cautious attitudes when returns are low and U_{bull} describes the attitudes when the market is booming. Both U_{bear} and U_{bull} are concave. However, due to switching the total utility function can be locally convex.

These utility functions are illustrated in figure 5.11 that shows the results of the sidewards market. We observe/estimate the market utility function that does not correspond to standard utility approaches because of the convex region. We propose to reconstruct this phenomenon by individual utility functions that consist of a bearish part and a bullish part. While the beginning of the bearish is fixed the bullish part starts at a switching point that characterizes the individual investor. By aggregating investors with different switching points we reconstruct the market utility function. We describe the aggregation in section 5.4.2 and estimate the distribution of switching points in section 5.4.3. In this way we explain the special form of the observed market utility functions.

5.4.2 Market Aggregation Mechanism

We consider the problem of aggregating individual utility functions to a representative market utility function. A simple approach to this problem is to identify the market utility function with an average of the individual utility functions. To this end one needs to specify the *observable* states of the world in the future by returns R and then find a weighted average of the utility functions for each state. If the importance of the investors is the same, then the weights are equal:

$$U(R) = \frac{1}{N} \sum_{i=1}^{N} U_i(R),$$

where N is the number of investors. The problem that arises in this case is that utility functions of different investors can not be summed up since they are incomparable.

Therefore, we propose an alternative aggregation technique. First we specify the *considered* or *subjective* states of the world given by future utility levels u and then aggregate the outlooks concerning the returns in the future R for each perceived state. For a *subjective* state described with the utility level U, such that

$$u = U_1(R_1) = U_2(R_2) = \ldots = U_N(R_N)$$

the aggregate estimate of the resulting return is

$$R_A(u) = \frac{1}{N} \sum_{i=1}^{N} U_i^{-1}(u) \tag{5.6}$$

if all investors have the same market power. The market utility function U_M resulting from this aggregation is given by the inverse R_A^{-1}.

In contrast to the naive approach described at the beginning of this section, this aggregation mechanism is consistent under transformations: If all individual utility functions are changed by the same transformation then the resulting market utility is also given by the transformation of the original aggregated utility. We consider the individual utility functions U_i and the resulting aggregate U_M. In addition, we consider the transformed individual utility functions $U_i^{\phi}(x) = \phi\{U_i(x)\}$ and the corresponding aggregate U_M^{ϕ} where ϕ is a transformation. Then the aggregation is consistent in the sense

that $U_M^\phi = \phi(U_M)$. This property can be seen from

$$\begin{aligned}(U_M^\phi)^{-1}(u) &= \frac{1}{N}\sum_{i=1}^{N}(U_i^\phi)^{-1}(u) \\ &= \frac{1}{N}\sum_{i=1}^{N}U_i^{-1}\{\phi^{-1}(u)\} \\ &= U_M^{-1}\{\phi^{-1}(u)\}\end{aligned}$$

The naive aggregation is not consistent in the above sense as the following example shows: We consider the two individual utility functions $U_1(x) = \sqrt{x}$ and $U_2(x) = \sqrt{x}/2$ under the logarithmic transformation $\phi = \log$. Then the naively aggregated utility is given by $U_M(x) = 3\sqrt{x}/4$. Hence, the transformed aggregated utility is $\phi\{U_M(x)\} = \log(3/4) + \log(x)/2$. But the aggregate of the transformed individual utility functions is

$$\begin{aligned}U_M^\phi(x) &= \frac{1}{2}\left\{\log(\sqrt{x}) + \log(\sqrt{x}/2)\right\} \\ &= \frac{1}{2}\log\left(\frac{1}{2}\right) + \log(x)/2.\end{aligned}$$

This implies that $U_M^\phi \neq \phi(U_M)$ in general.

This described aggregation approach can be generalized in two ways: If the individual investors have different market power then we use the corresponding weights w_i in the aggregation (5.6) instead of the uniform weights. As the number of market participants is in general big and unknown it is better to use a continuous density f instead of the discrete distributions given by the weights w_i. These generalizations lead to the following aggregation

$$R_A(u) = \int U^{-1}(\cdot, z)(u) f(z) dz$$

where $U(\cdot, z)$ is the utility function of investor z. We assume in the following that the investors have utility function of the form described in section 5.4.1. In the next section we estimate the distribution of the investors who are parametrized by z.

5.4.3 Estimation of the Distribution of Switching Points

Using the described aggregation procedure, we consider now the problem of replicating the market utility by aggregating individual utility functions. To

this end, we choose the parametric utility functions $U(\cdot, z)$ described in 5.4.1 and try to recover with them the market utility U_M. We do not consider directly the utility functions but minimize instead the distance between the inverse functions:

$$\min_f \| \int U^{-1}(\cdot, z) f(z) dz - U_M^{-1} \|_{L^2(\tilde{P})} \tag{5.7}$$

where $\tilde{P}$ is image measure of the historical measure P on the returns under the transformation U_M. As the historical measure has the density p the transformation theorem for densities implies that $\tilde{P}$ has the density

$$\tilde{p}(u) = p\{U_M^{-1}(u)\} / U_M'\{U_M^{-1}(u)\}.$$

With this density the functional to be minimized in problem (5.7) can be stated as

$$\begin{aligned}
&\int \left(\int U^{-1}(u, z) f(z) dz - U_M^{-1}(u) \right)^2 \tilde{p}(u) \, du \\
=&\int \left(\int U^{-1}(u, z) f(z) dz - U_M^{-1}(u) \right)^2 p\{U_M^{-1}(u)\} / U_M'\{U_M^{-1}(u)\} \, du \\
=&\int \left(\int U^{-1}(u, z) f(z) dz - U_M^{-1}(u) \right)^2 p\{U_M^{-1}(u)\} (U_M^{-1})'(u) \, du
\end{aligned}$$

because the derivative of the inverse is given by $(g^{-1})'(y) = 1/g'\{g^{-1}(y)\}$. Moreover, we can apply integration by substitution to simplify this expression further

$$\begin{aligned}
&\int \left(\int U^{-1}(u, z) f(z) dz - U_M^{-1}(u) \right)^2 p\{U_M^{-1}(u)\} (U_M^{-1})'(u) \, du \\
=&\int \left(\int U^{-1}\{U_M(x), z\} f(z) dz - x \right)^2 p(x) \, dx.
\end{aligned}$$

For replicating the market utility by minimizing (5.7) we observe first that we have samples of the historical distribution with density p. Hence, we can replace the outer integral by the empirical expectation and the minimization problem can be restated as

$$\min_f \frac{1}{n} \sum_{i=1}^{n} \left(\int g\{U_M(x_i), z\} f(z) dz - x_i \right)^2$$

where $x_1 \ldots, x_n$ are the samples from the historical distribution and $g = U^{-1}$.

Replacing the density f by a histogram $f(z) = \sum_{j=1}^{J} \theta_j I_{B_j}(z)$ with bins B_j the problem is transformed into

$$\min_{\theta_j} \frac{1}{n} \sum_{i=1}^{n} \left\{ \sum_{j=1}^{J} \tilde{g}(i,j)\theta_j - x_i \right\}^2$$

where $\tilde{g}(i,j) = \int_{B_j} g\{U_M(x_i), z\} dz$.

Hence, the distribution of switching points can be estimated by solving the quadratic optimization problem

$$\begin{aligned} \min_{\theta_j} \frac{1}{n} \sum_{i=1}^{n} \left\{ \sum_{j=1}^{J} \tilde{g}(i,j)\theta_j - x_i \right\}^2, \\ \text{s.t.} \qquad \theta_j \geq 0, \\ \sum_{j=1}^{J} \theta_j |B_j| = 1. \end{aligned}$$

Such quadratic optimization problems are well known and their solutions can be obtained using standard techniques, see e.g. Mehrotra (1992) or Wright (1998).

We present in figure 5.12 the estimated distribution of switching points of the sidewards market. For this picture we have replaced the density f by a histogram with bins of width 0.01. We checked the robustness of our results by considering also smaller and bigger bandwidths. Smaller bandwidths led to rather ragged densities and bigger bandwidths deleted too much structure. The density presented in figure 5.12 shows that the distribution is concentrated on a small region. Moreover, it is left-skewed. As the switching points lie in the area of 5% to 10% returns the investors do not seem to use the riskless return as reference. Rather they use a higer returns that could correspond to the expected return of the DAX. The long right tail of the distribution means that there are some relatively pessimistic investors that switch to the bullish regime only for high returns. The switching points are all bigger than the riskless return, hence all investors want in a sidewards situation more than only positive returns in order to show bullish attitudes. The corresponding distibutions of the bullish and bearish markets of 24/03/2000 and 30/07/2002 show similar characteristics only the region where the densities are concentrated are bigger.

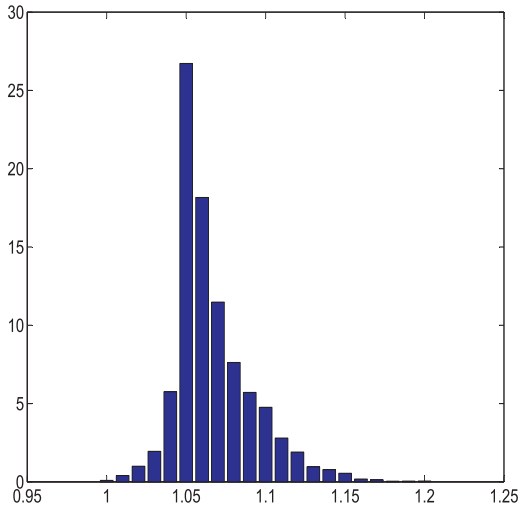

Figure 5.12: The density of the distribution of the reference points in the sidewards market of 24/03/2000.

5.5 Conclusion

We have analyzed in this section empirical pricing kernels in three market regimes using data on the German stock index and options on this index. In a bullish, a bearish and a neutral market regime we estimate the pricing kernel and derive the corresponding utility functions and relative risk aversions.

In the sidewards market of June 2004, the market investor is risk seeking in a small region around the riskless return but risk aversion increases fast for high absolute returns. In the bullish market of March 2000, the investor is on the other hand never risk seeking while she become more risk seeking in the bearish market of July 2002. Before the stock market crash in 1987 European options did not show the smile and the Black-Scholes model captured the data quite well. Hence, utility functions could be estimated at that times by power utility functions with a constant positive risk aversion. Our analysis shows that this simple structure does not hold anymore and discusses different structures corresponding to different market regimes.

The empirical pricing kernels of all market regimes demonstrate that the corresponding utility functions do not correspond to standard specifications of utility functions including Kahneman and Tversky (1979). The observed utility functions correspond the most to the general utility functions of Friedman and Savage (1948). We propose a parametric specification of these functions, estimate it and explain the observed market utility function by aggregating individual utility functions. In this way, we can estimate a dis-

tribution of individual investors.

The proposed aggregation mechanism is based on homogeneous investors in the sense that the differ only by a switching points. Future research can analyze how nonlinear aggregation procedures could be applied to heterogeneous investors.

Bibliography

Ait-Sahalia, Y. and Lo, A. (1998). Nonparametric estimation of state-price densities implicit in financial asset prices. *Journal of Finance*, 53(2).

Ait-Sahalia, Y. and Lo, A. (2000). Nonparametric risk-management and implied risk aversion. *Journal of Econometrics*, 94(9).

Albrecher, H., Mayer, P., Schoutens, W., and Tistaert, J. (2007). The little heston trap. *Wilmott*, pages 83–92.

Barone-Adesi, G., Engle, R., and Mancini, L. (2004). Garch options in incomplete markets. working paper, University of Lugano.

Bates, D. S. (1996). Jumps and stochastic volatility: Exchange rate processes implicit in deutsche mark options. *Review of Financial Studies*, 9(1):69–107.

Bergomi, L. (2004). Smile dynamics. *Risk*, 17(9).

Bergomi, L. (2005). Smile dynamics 2. *Risk*, 18(10).

Bernoulli, D. (1956). Exposition of a new theory on the measurement of risk. *Econometrica*, 22:23–36.

Billingsley, P. (1995). *Probability and Measure*. Wiley-Interscience.

Black, F. and Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, 81:637–659.

Breeden, D. and Litzenberger, R. (1978). Prices of state-contingent claims implicit in option prices. *Journal of business*, 51:621–651.

Broadie, M. and Kaya, O. (2006). Exact simulation of stochastic volatility and other affine jump diffusion processes. *Operations Research*, 54(2):217–231.

Bühler, H. (2006). Consistent variance curve models. *Finance and Stochastics*, 10(2):178–203.

Carr, P., Geman, H., Madan, D., and Yor, M. (2002). The fine structure of asset returns: An empirical investigation. *Journal of business*, 75.

Carr, P. and Madan, D. (1998). Towards a theory of volatility trading. *Volatility, RISK Publications*.

Carr, P. and Madan, D. (1999). Option valuation using the fast fourier transform. *Journal of Computational Finance*, 2:61–73.

Chernov, M. (2000). Essays in financial econometrics. Phd thesis, Pennsylvania State University.

Chernov, M. (2003). Empirical reverse engineering of the pricing kernel. *Journal of Econometrics*, 116:329–364.

Chourdakis, K. (2005). Option pricing using the fractional fft. *Journal of Computational Finance*, 8(2):1–18.

Cizek, P., Härdle, W., and Weron, R. (2005). *Statistical Tools in Finance and Insurance*. Springer, Berlin.

Cochrane, J. (2001). *Asset Pricing*. Princeton University Press.

Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues.

Cont, R. (2005). Model uncertainty and its impact on the pricing of derivative instruments. *Mathematical Finance*, 16:519–542.

Cont, R. and da Fonseca, J. (2002). Dynamics of implied volatility surfaces. *Quantitative Finance*, 2(1):45–60.

Cont, R. and Tankov, P. (2004a). *Financial Modelling with Jump Processes*. Chapman & Hall.

Cont, R. and Tankov, P. (2004b). Nonparametric calibration of jump-diffusion option pricing models. *Journal of Computational Finance*, 7(3):1–49.

Cox, J. (1996). The constant elasticity of variance option pricing model. *Journal of Portfolio Management*, 23(1):5–17.

Deelstra, G. and Delbaen, F. (1998). Convergence of discretized stochastic (interest rate) processes with stochastic drift term. *Applied stochastic models and data analysis*, 14(1):77–84.

Demeterfi, K., Derman, E., Kamal, M., and Zou, J. (1999). More than you ever wanted to know about volatility swaps. Quantitative strategies research notes, Goldman Sachs.

Derman, E. and Kani, I. (1994a). Riding on the smile. *Risk*, 7(2).

Derman, E. and Kani, I. (1994b). The volatility smile and its implied tree. Technical report, Goldman Sachs.

Diebold, F. and Li, C. (2006). Forecasting the term structure of government bonds yields. *Journal of Econometrics*, 130.

Duffie, D., Pan, J., and Singleton, K. (2000). Transform analysis and asset pricing for affine jump-diffusions. *Econometrica*, 68(6):1343–1376.

Duffy, D. (2006). *Finite difference methods in financial engineering*. John Wiley & Sons.

Dupire, B. (1994). Pricing with a smile. *Risk*, 7:327–343.

Fengler, M. (2005). *Semiparametric Modeling of Implied Volatility*. Springer, Berlin.

Figlewski, S. (1989). Options arbitrage in imperfect markets. *The Journal of Finance*, 44(5):1289–1311.

Franke, J., Härdle, W., and Hafner, C. (2004). *Statistics of Financial Markets*. Springer Verlag, Berlin.

Friedman, M. and Savage, L. P. (1948). The utility analysis of choices involving risk. *Journal of Political Economy*, 56:279–304.

Gasser, T., Kneip, A., and Kohler, W. (1991). A flexible and fast method for automatic smoothing. *J. Amer. Statist. Assoc.*, 86.

Glasserman, P. (2004). *Monte Carlo Methods in Financial Engineering*. Springer, New York.

Hagan, P., Kumar, D., A, L., and D, W. (2002). Managing smile risk. *Wilmott Magazine*.

Härdle, W., Müller, M., Sperlich, S., and Werwatz, A. (2004). *Nonparametric and semiparametric models*. Springer, Heidelberg.

Härdle, W. and Simar, L. (2003). *Applied Multivariate Statistical Analysis.* Springer, Berlin.

Harrison, M. and Pliska, S. (1981). Martingales and stochastic integrals in the theory of continuous trading. *Stochastic Processes and their Applications*, 11:215–260.

Harvey, A. (1989). *Forecasting, Structural time series models and Kalman filtering*. Cambridge University Press.

Heston, S. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies*, 6(2):327–343.

Heston, S. and Nandi, S. (2000). A clsed form garch option pricing model. *Review of Financial Studies*, 13:585–625.

Hull, J. and White, A. (1987). The pricing of options on assets with stochastic volatilities. *Journal of Finance*, 42:281–300.

Jackwerth, J. (2000). Recovering risk aversion from option prices and realized returns. *Review of Financial Studies*, 13(2):433–451.

Jackwerth, J. and Rubinstein, M. (1996). Recovering probability distributions from option prices. *Journal of Finance*, 51(5):1611–1631.

Kahl, C. and Jäckel, P. (2005). Not-so-complex logarithms in the heston model. *Wilmott*, pages 94–103.

Kahneman, D. and Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47:263–291.

Ledoit, O., Santa-Clara, P., and Yan, S. (2002). Relative pricing of options with stochastic volatility. working paper, UCLA.

Lord, R., Koekkoek, R., and van Dijk, D. (2006). A comparison of biased simulation schemes for stochastic volatility models. working paper, Erasmus University Rotterdam.

Madan, D. and Seneta, E. (1990). The variance gamma process for share market returns. *Journal of Business*, 63:511–524.

Mehrotra, S. (1992). On the implementation of a primal-dual interior point method. *SIAM Journal on Optimization*, 2(4):575–601.

Merton, R. (1976). Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics*, 3:125–183.

Merton, R. C. (1971). Optimum consumption and portfolio rules in a continuous-time model. *Journal of Economic Theory*, 3:373–413.

Merton, R. C. (1973). The theory of rational option pricing. *Bell Journal of Economics and Management Science*, 4:124–144.

Mikhailov, S. and Nögel, U. (2003). Heston's stochastic volatility model. implementation, calibration and some extensions. *Wilmott*, pages 74–94.

Nelson, C. and Siegel, A. (1987). Parsimonious modeling of the yield curve. *Journal of Business*, 60:473–89.

Neuberger, A. (1992). Volatility trading. *working paper*.

Protter, P. (2004). *Stochastic Integration and Differential Equations*. Springer, second edition.

Rosenberg, J. and Engle, R. (2002). Empirical pricing kernels. *Journal of Financial Economics*, 64(7):341–372.

Rubinstein, M. (1994). Implied binomial trees. *Journal of Finance*, 69:771–818.

Schöbel, R. and Zhu, J. (1999). Stochastic volatility with an ornstein uhlenbeck process: An extension. *European Finance Review*, 3.

Schönbucher, P. (1999). A market model for stochastic implied volatility. *Phil. Trans. Royal Soc. Series A*, 357:2071–2092.

Schoutens, W., Simons, E., and Tistaert, J. (2004). A perfect calibration! now what? *Wilmott*, pages 66–78.

Scott, L. (1987). Option pricing when the variance changes randomly: theory, estimation and an application. *Journal of Financial and Quantitative Analysis*, 2(4):419–438.

Silverman, B. (1986). *Density Estimation*. Chapman and Hall, London.

Stein, E. and Stein, J. (1991). Stock price distribution with stochastic volatility: An analytic approach. *Review of Financial Studies*, 4:727–752.

Storn, R. and Price, K. (1997). Differential evolution - a simple and efficient heuristic for global optimization over continuous space. *Journal of Global Optimization*, 11:341–359.

Topper, J. (2005). *Financial engineering with finite elements*. Wiley, New York, NY.

von Neumann, J. and Morgenstern, O. (1944). *The Theory of Games and Economic Behavior*. Princeton University Press.

Wright, S. (1998). Primal-dual interior-point methods. *Mathematics of Computation*, 67(222):867–870.

Selbständigkeitserklärung

Ich bezeuge durch meine Unterschrift, dass meine Angaben über die bei der Abfassung benutzten Hilfsmittel, über die mir zuteil gewordene Hilfe sowie über frühere Begutachtungen meiner Dissertation in jeder Hinsicht der Wahrheit entsprechen.